

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Узунов Федор Владимирович

Должность: Ректор

Дата подписания: 19.06.2026 18:16:50

Уникальный программный ключ:

fd935d10451b860e912264c037866448452bfdb603f94388008a29877a6bcbf5

**АВТНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ  
«ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ»**

**«УНИВЕРСИТЕТ ЭКОНОМИКИ И УПРАВЛЕНИЯ»**

**Факультет экономики, управления и юриспруденции**

**Кафедра управления и бизнес-информатики**

**УТВЕРЖДАЮ**

Проректор по учебно-методической работе

 / Г.П. Узунова

«02» февраля 2026 г.



**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ**

**ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ**

*Направление подготовки*

**09.03.01 Информатика и вычислительная техника**

*Профиль:* специалист по компьютерным системам

Квалификация выпускника: бакалавр

Для всех  
форм обучения

Симферополь, 2026 г.

## 1. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ

Средства оценивания в ходе текущего контроля:

- устные опросы в ходе семинарских занятий;
- рефераты;
- тестирование;
- практические задания, выполняемые в ходе семинарского (практического) занятия или рекомендуемые для самостоятельной работы.

Коды компетенции	Результаты освоения ОПОП	Перечень планируемых результатов обучения по дисциплине
ОПК-5	Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.1. Знать: основы математики, физики, вычислительной техники и программирования. ОПК-1.2. Уметь: решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования. ОПК-1.3. Владеть: навыками теоретического и экспериментального исследования объектов профессиональной деятельности

### 1.1 Вопросы к текущему контролю

- 1 Что такое данные?
- 2 Что такое информация?
- 3 Что такое интеллектуальный анализ данных?
- 4 Какие задачи решает Data Mining?
- 5 Что такое набор данных (dataset)?
- 6 Что такое признак (feature)?
- 7 Что такое объект наблюдения?
- 8 Что такое целевая переменная?
- 9 Что такое выборка?
- 10 Чем отличается обучающая выборка от тестовой?
- 11 Что такое пропущенные данные?
- 12 Как можно обработать пропуски?
- 13 Что такое выбросы?
- 14 Как обнаружить выбросы?
- 15 Что такое шум в данных?
- 16 Зачем нужна очистка данных?
- 17 Что такое нормализация?
- 18 Что такое стандартизация?
- 19 Зачем масштабировать данные?
- 20 Что такое кодирование категориальных признаков?
- 21 Что такое классификация?
- 22 Приведите пример задачи классификации.
- 23 Что такое регрессия?
- 24 Приведите пример задачи регрессии.
- 25 Что такое кластеризация?

- 26 Чем кластеризация отличается от классификации?
- 27 Что такое обучение с учителем?
- 28 Что такое обучение без учителя?
- 29 Что такое переобучение?
- 30 Что такое недообучение?
- 31 Что такое kNN?
- 32 Как работает kNN?
- 33 Что такое расстояние в kNN?
- 34 Что такое наивный байес?
- 35 Что такое вероятность в контексте Байеса?
- 36 Что такое дерево решений?
- 37 Что такое узел дерева решений?
- 38 Что такое лист дерева решений?
- 39 Что такое разбиение в дереве решений?
- 40 Что такое критерий разбиения?
- 41 Что такое энтропия?
- 42 Что такое информационный выигрыш?
- 43 Что такое случайный лес?
- 44 Что такое ансамбль моделей?
- 45 Что такое бустинг?
- 46 Что такое бэггинг?
- 47 В чем разница между бустингом и бэггингом?
- 48 Что такое линейная модель?
- 49 Что такое коэффициенты модели?
- 50 Что такое функция потерь?
- 51 Что такое логистическая регрессия?
- 52 Для чего используется логистическая регрессия?
- 53 Что такое сигмоидная функция?
- 54 Что такое вероятность класса?
- 55 Что такое метрика качества модели?
- 56 Что такое accuracy?
- 57 Что такое precision?
- 58 Что такое recall?
- 59 Что такое F1-мера?
- 60 Зачем оценивать качество модели?

## **1.2 Темы рефератов:**

1. История развития интеллектуального анализа данных.
2. Основные этапы процесса KDD.
3. Роль Data Mining в современных информационных системах.
4. Классификация методов интеллектуального анализа данных.
5. Предобработка данных как ключевой этап анализа.
6. Методы очистки и подготовки данных.
7. Проблема качества данных в аналитических системах.
8. Работа с пропущенными значениями в больших данных.
9. Обнаружение и обработка выбросов.
10. Масштабирование и нормализация данных.
11. Методы классификации в Data Mining.
12. Алгоритм k-ближайших соседей и его применение.
13. Деревья решений: принципы и особенности.
14. Ансамблевые методы машинного обучения.
15. Случайный лес: теория и практика.
16. Градиентный бустинг в задачах анализа данных.

17. Наивный байесовский классификатор.
18. Логистическая регрессия в задачах классификации.
19. Линейная регрессия и её применение.
20. Функции потерь в машинном обучении.
21. Методы кластеризации данных.
22. Алгоритм k-means: особенности и ограничения.
23. Иерархическая кластеризация.
24. Алгоритм DBSCAN и анализ плотности.
25. Методы снижения размерности данных.
26. Метод главных компонент (PCA).
27. Визуализация многомерных данных.
28. Методы отбора признаков (feature selection).
29. Оценка качества моделей машинного обучения.
30. Применение интеллектуального анализа данных в реальных задачах (медицина, финансы, маркетинг, безопасность).

### 1.3 Тестовые задания

#### 1. Что такое интеллектуальный анализ данных?

- а) Процесс извлечения знаний из данных (*Правильный ответ: а*)
- б) Хранение данных в базе
- в) Передача данных по сети
- г) Шифрование данных

#### 2. Какая задача относится к классификации?

- а) Прогноз температуры
- б) Сжатие данных
- в) Разделение писем на спам/не спам (*Правильный ответ: в*)
- г) Шифрование

#### 3. Что такое кластеризация?

- а) Удаление данных
- б) Кодирование данных
- в) Разбиение данных на группы без учителя (*Правильный ответ: в*)
- г) Передача данных

#### 4. Какой алгоритм относится к классификации?

- а) k-means
- б) PCA
- в) kNN (*Правильный ответ: в*)
- г) DBSCAN

#### 5. Что такое переобучение модели?

- а) Увеличение данных
- б) Слишком хорошее запоминание обучающей выборки (*Правильный ответ: б*)
- в) Удаление признаков
- г) Снижение размерности

#### 6. Что измеряет ассигасу?

- а) Среднюю ошибку
- б) Размер модели
- в) Долю верных предсказаний (*Правильный ответ: в*)
- г) Время обучения

#### 7. Что такое признак?

- а) Результат модели
- б) Свойство объекта данных (*Правильный ответ: б*)
- в) Ошибка
- г) Алгоритм

#### 8. Какой метод относится к регрессии?

- а) Логистическая регрессия
- б) DBSCAN
- в) Линейная регрессия (*Правильный ответ: в*)
- г) k-means

**9. Что делает PCA?**

- а) Классифицирует данные
- б) Удаляет выбросы
- в) Снижает размерность данных (*Правильный ответ: в*)
- г) Обучает модель

**10. Что такое обучающая выборка?**

- а) Данные для проверки
- б) Ошибки
- в) Данные для обучения модели (*Правильный ответ: в*)
- г) Результаты

**11. Какие задачи относятся к Data Mining?**

- а) Классификация
- б) Кластеризация
- в) Архивация файлов
- г) Регрессия

**Ответ:** а, б, г.

**12. Что относится к предобработке данных?**

- а) Очистка данных
- б) Нормализация
- в) Обучение модели
- г) Обработка пропусков

**Ответ:** а, б, г.

**13. Какие модели относятся к классификации?**

- а) kNN
- б) PCA
- в) дерево решений
- г) Наивный Байес

**Ответ:** а, в, г.

**14. Что может быть проблемой в данных?**

- а) Выбросы
- б) Пропуски
- в) Шифрование
- г) Шум

**Ответ:** а, б, г.

**15. Какие методы относятся к кластеризации?**

- а) k-means
- б) DBSCAN
- в) SVM
- г) Иерархическая кластеризация

**Ответ:** а, б, г.

**16. Какие метрики качества используются?**

- а) Accuracy
- б) Precision

- в) Recall
- г) Frequency

**Ответ:** а, б, в.

**17. Какие методы являются ансамблевыми?**

- а) Случайный лес
- б) Бустинг
- в) kNN
- г) Бэггинг

**Ответ:** а, б, г.

**18. Какие действия характерны для DDoS-атаки?**

- а) Перегрузка сервера запросами
- б) Удаление файлов с сервера
- в) Использование ботнета
- г) Шифрование данных

**Ответ:** а, б, г.

**19. Какие методы относятся к снижению размерности?**

- а) PCA
- б) t-SNE
- в) kNN
- г) LDA

**Ответ:** а, б, г.

**20. Что влияет на качество модели?**

- а) Качество данных
- б) Выбор признаков
- в) Размер монитора
- г) Алгоритм обучения

**Ответ:** а, б, г.

**21. Установить соответствие:**

1. kNN	А) Снижение размерности
2. PCA	Б) Кластеризация
3. k-means	В) Классификация
4. Логистическая регрессия	Г) Регрессия/классификация на основе вероятности

**22. Установить соответствие:**

1. Линейная регрессия	А) С учителем
2. Дерево решений	Б) Без учителя
3. DBSCAN	В) Полу-обучение
4. k-means	Г) Ансамблевый метод

**23. Установить соответствие:**

1. Признак	А) Отсутствующее значение
2. Выброс	Б) Выброс
3. Пропуск	В) Пропуск
4. Объект данных	Г) Объект данных

**24. Установить соответствие:**

1. Логистическая регрессия	А) Задача классификации
2. Линейная регрессия	Б) Задача кластеризации
3. PCA	В) Задача снижения размерности
4. k-means	Г) Задача предсказания числового значения

**25. Установить соответствие:**

1. Accuracy	А) Доля верных предсказаний
2. Precision	Б) Полнота
3. Recall	В) Баланс точности и полноты
4. F1-мера	Г) Точность положительных предсказаний

**26. Установить соответствие:**

1. DBSCAN	А) Класс кластеризации
2. kNN	Б) Класс классификации
3. PCA	В) Класс снижения размерности
4. Случайный лес	Г) Класс ансамблевого метода

**27. Установить соответствие:**

1. for	А) Выполняется минимум 1 раз
2. while	Б) Проверка в начале
3. do while	В) Нет условия выхода
4. бесконечный цикл	Г) Цикл с параметром

**28. Установить соответствие:**

1. Пропуски в данных	А) Нормализация
2. Выбросы	Б) Удаление/замена значений
3. Разные масштабы признаков	В) Очистка данных

4. Шум	Г) Фильтрация/сглаживание
--------	---------------------------

**29. Установить соответствие:**

1. Классификация	А) Без учителя
2. Кластеризация	Б) С учителем
3. Регрессия	В) Предсказание чисел
4. Ассоциация	Г) Поиск закономерности между объектами

**30. Установить соответствие:**

1. Overfitting	А) Слишком простая модель
2. Underfitting	Б) Переобучение
3. Noise	В) Систематическая ошибка модели
4. Bias	Г) Случайные искажения данных

**1.4 Практические задания**

**Практическое задание № 1.**

Основы работы с numpy.

Задача: Требуется сложить два целых числа А и В.

Везде значение  $n+10$  — это Ваш порядковый номер в списке группы + 10

1. Напечатать версию и конфигурацию numpy
2. Создать вектор размерности  $n$ , заполненный случайными числами
3. Создать вектор размерности  $n$ , заполненный нулями
4. Создать вектор размерности  $n$ , заполненный единицами
5. Создать вектор размерности  $n$ . Если Ваш порядковый номер — четное число, то половина вектора должна быть заполнена единицами, а другая половина нулями. Если Ваш порядковый номер — нечетное число, например, 15, то первые 7 элементов вектора должны быть единицами, 8-й — Ваш порядковый номер, оставшиеся 7 элементов — нули.
6. Создать вектор размерности  $n$ , заполненный Вашим порядковым номером + 7.5
7. Создать вектор размерности  $n$ , заполненный единицами, но для элемента под номером, равным Ваш порядковый номер+какое-то число (не выходя за диапазон вектора) следует установить значение -15, а для элемента под номером, равным Ваш порядковый номер-какое-то число (не выходя за диапазон вектора) следует установить значение +25
8. Создать вектор с диапазоном — от числа, равного Вашему порядковому номеру до этого числа +15
9. Создать вектор размерности  $n$ , заполненный случайными числами и развернуть этот вектор
10. Создать матрицу размерности  $n \times n$  со случайными значениями
11. Найти индексы ненулевых элементов в векторе размерности  $n$  со случайными значениями
12. Вычислить количество нулевых элементов в векторе размерности  $n$  со случайными значениями
13. Найти минимальное, максимальное и среднее значение элементов в матрице размерности  $3 \times 3$  со случайными значениями

14. Создать матрицу размерности  $n \times n$  с 0 внутри и 1 на границах
15. Создать матрицу размерности  $n \times n$  с  $1, 2, 3, \dots, n-1$  под диагональю
16. Создать матрицу размерности  $n \times n$  с  $n-1, \dots, 2, 1$  над диагональю
17. Создать матрицу размерности  $8 \times n \times 8 \times n$  и заполнить её единицами в шахматном порядке
18. Создать матрицу размерности  $8 \times n \times 8 \times n$  и заполнить её единицами в шахматном порядке, используя `tile`
19. Перемножить любые матрицы размерности  $7 \times n$  и  $n \times 5$
20. Создать вектор размерности  $n$ . Поменять знак у элементов, значения которых между  $n$  и  $3 \times n$

### Практическое задание № 2.

Основы работы с pandas (объект Series).

1. Получить элементы объекта Series A, которых нет в объекте Series B
2. Получить не пересекающиеся элементы в двух объектах Series
3. Заменить все элементы объекта Series на "Other", кроме двух наиболее часто встречающихся
4. Найти индексы объекта Series кратные 3
5. Объединить два объекта Series вертикально и горизонтально
6. Рассчитать количество символов в объекте Series
7. Отобразить элементы объекта Series, которые содержат не менее двух гласных
8. Найти евклидово расстояние между двумя объектами Series

### Практическое задание № 3.

Основы работы с pandas (объект DataFrame).

1. Получить наименование столбцов и сумму пропущенных значений DataFrame (брать файл `titanic.csv`)
2. Изменить позиции столбцов объекта DataFrame (создать свой объект DataFrame)
  - # 1. Поменять местами столбцы 'a' и 'c'
  - # 2. Написать функцию, которая меняет столбцы местами
  - # 3. Сортировать столбцы по наименованию
3. Отобразить DataFrame по нескольким условиям поля "Age" и "Sex" (пол - мужской, возраст - больше 30) (брать файл `titanic.csv`)
4. Отобразить DataFrame по полю "Age" (между 30 и 40) (брать файл `titanic.csv`)
5. Отобразить каждую 20-ую строку и столбцы ['Age', 'Name', 'Survived'] (`titanic.csv`)

создать объект DataFrame на основе своего csv файла с данными о студентах.

- #1. Отобразить студентов со средним баллом выше 4.5
- #2. Отобразить студентов, у которых id зачетки оканчивается на 5
- #3. Отобразить студентов, у которых фамилия начинается с буквы "К"

### Практическое задание № 4.

Основы работы с matplotlib.

1. Построить график функции  $y = \sin(x)$
2. Нарисуйте круговую диаграмму и гистограмму, демонстрирующие диаметры планет (справочная информация: <http://www.webteka.com/solarsystem-diameters-planets/>)
3. Прочитать файл `hubble.txt` с данными измерений космического телескопа Хаббл и нарисовать диаграмму зависимости скорости убегания от объекта измерения

### Практическое задание № 5.

Обработка текста. Токенизация. Лемматизация.

Возьмите текст о Лондоне на английском языке из википедии. Пример: London is the capital and largest city of England and of the United Kingdom. Standing on the River

Thames in the south-east of England, at the head of its 50-mile (80 km) estuary leading to the North Sea, London has been a major settlement for two millennia. Londinium was founded by the Romans.[9] The City of London, London's ancient core – an area of just 1.12 square miles (2.9 km<sup>2</sup>) and colloquially known as the Square Mile – retains boundaries that closely follow its medieval limits.[10][11][12][13][14][note 1] The City of Westminster is also an Inner London borough holding city status. Greater London is governed by the Mayor of London and the London Assembly.

Напишите сценарий для очистки текста от ссылок на источники в квадратных скобках. Используйте регулярные выражения

Выполните токенизацию предложений и слов в собственном тексте

Выведите список имен существительных, прилагательных и глаголов из собственного текста

Для собственного русскоязычного текста вывести все существительные в винительном падеже

Выполните лемматизацию и выведите все начальные формы слов из собственного текста

Выведите список стоп-слов из собственного текста. Расширьте список стоп-слов собственным словарем и удалите стоп-слова из текста согласно расширенному списку.

#### **Практическое задание № 6.**

Обработка текста. Выделение сущностей.

Выведите список именованных сущностей и группы слов, относящихся к каждой выведенной именованной сущности. Например, London, England, United Kingdom — Geographic Entity

Из собственного текста вывести пары Прилагательное+Существительное, имеющие негативную тональность (сентимент).

Взять собственный текст и организовать голосование алгоритмов на словарях VADER, Afinn, Textblob для принятия решений о тональности предложений по большинству голосов.

Возьмите 15 отзывов с amazon (5 позитивных отзывов, 5 негативных отзывов и 5 нейтральных отзывов). Оцените тональность отдельно по каждому из алгоритмов VADER, Afinn, Textblob. Сделайте вывод, какой алгоритм точнее.

#### **Практическое задание № 7.**

Обработка текста. Поиск и изменение текста по критерию.

1. Дан текст. Вывести слово с наибольшим количеством гласных
2. Дан текст. Найти количество существительных из разных категорий.
3. Дан текст. Заменить все глаголы на глаголы в Present Continuous (добавить ing и правильный глагол to be)
4. Дан текст. Напечатать все биграммы текста и выявить 5 биграмм с наибольшей частотой
5. Дан текст. Удалить все прилагательные из текста.
6. Дан текст. Определить гендерную принадлежность текста частотным методом

#### **Практическое задание № 8.**

Введение в PIL/Pillow.

Задание: создать красивое абстрактное изображение с использованием одной из арифметических операций над изображениями

#### **Практическое задание № 9.**

Обработка изображений с OpenCV.

Загрузить изображение в OpenCV и изменить его.

#### **Практическое задание № 10.**

Цветовая сегментация. Выполните цветовую сегментацию, взяв за основу фото любого животного.

## 2. КРИТЕРИИ ОЦЕНИВАНИЯ ПРИ ПРОВЕДЕНИИ ТЕКУЩЕГО КОНТРОЛЯ

Вид контроля	Наименование работы	Наименование оценочных средств	Шкала оценивания
Текущий контроль	<ul style="list-style-type: none"> <li>- Вопросы для обсуждения на занятиях;</li> <li>- Устные опросы по ранее изученному материалу;</li> <li>- Письменные работы: рефераты, тестовые задания;</li> <li>- Практические задания;</li> <li>- Рефераты и доклады по темам (вопросам), вынесенным на самостоятельную работу.</li> </ul>	Оценка выступлений на практическом (семинарском) занятии, проверка заданий и аудиторных работ, устный опрос, оценивание докладов, рефератов	отлично хорошо удовлетворительно неудовлетворительно

### Критерии оценивания устных ответов обучающихся

Шкала оценивания	Характеристика оценивания
отлично	оценивается ответ, который показывает прочные знания основных процессов изучаемой предметной области, отличается глубиной и полнотой раскрытия темы; владение терминологическим аппаратом; умение объяснять сущность, явлений, процессов, событий, делать выводы и обобщения, давать аргументированные ответы, приводить примеры; свободное владение монологической речью, логичность и последовательность ответа.
хорошо	оценивается ответ, обнаруживающий прочные знания основных процессов изучаемой предметной области, отличается глубиной и полнотой раскрытия темы; владение терминологическим аппаратом; умение объяснять сущность, явлений, процессов, событий, делать выводы и обобщения, давать аргументированные ответы, приводить примеры; свободное владение монологической речью, логичность и последовательность ответа. Однако допускается одна - две неточности в ответе.
удовлетворительно	оценивается ответ, свидетельствующий в основном о знании процессов изучаемой предметной области, отличающийся недостаточной глубиной и полнотой раскрытия темы; знанием основных вопросов теории; слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры; недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа.
неудовлетворительно	оценивается ответ, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы; незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов; неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Допускаются серьезные ошибки в содержании ответа.

### Критерии оценивания работы обучающихся на практических и семинарских занятиях

Шкала оценивания	Показатели	Критерии
Отлично	<ol style="list-style-type: none"> <li>1. Полнота выполнения практического и тестового задания (полнота ответа);</li> <li>2. Своевременность выполнения задания;</li> <li>3. Последовательность и рациональность выполнения практического задания (логичность и четкость ответа);</li> </ol>	<p>Задание решено самостоятельно. При этом составлен правильный алгоритм решения задания, в логических рассуждениях, в выборе формул и решении нет ошибок, получен верный ответ, задание решено рациональным способом.</p> <p>Дан правильный и исчерпывающий ответ на поставленные теоретические и тестовые вопросы, в которых обучающийся показал всестороннее системное знание программного материала, усвоение основной и дополнительной литературы, четкое владение понятийным аппаратом.</p>
Хорошо	<ol style="list-style-type: none"> <li>4. Правильность ответов на вопросы;</li> <li>5. Самостоятельность решения (владение дополнительным материалом);</li> <li>6. Знание нормативно-законодательной базы и терминологии курса</li> </ol>	<p>Задание решено с помощью преподавателя. При этом составлен правильный алгоритм решения задания, в логическом рассуждении и решении нет существенных ошибок; правильно сделан выбор формул для решения; есть объяснение решения, но задание решено нерациональным способом или допущено не более двух несущественных ошибок, получен верный ответ.</p> <p>На поставленные теоретические и тестовые вопросы, при которых обучающийся показал достаточный уровень знаний основного программного материала: освоение информации лекционного курса и учебных пособий, овладение понятийным аппаратом, методикой исследований при попытке анализа различных ситуаций.</p>
Удовлетворительно		<p>Задание решено с подсказками преподавателя. Задание решено в общем виде.</p> <p>Обучающийся показал средний уровень знаний основного программного материала, но не мог убедительно аргументировать свой ответ, ошибся в использовании понятийного аппарата, показал недостаточные знания литературных источников.</p>
Неудовлетворительно		<p>Задание не решено.</p> <p>Обучающийся продемонстрировал значительные пробелы в знаниях основного программного материала, не аргументировал свой ответ, показал неудовлетворительные знания понятийного аппарата и специальной литературы.</p>

### Критерии оценивания рефератов

Средство контроля	Критерии оценивания	Шкала оценивания
Реферат	Выполнены все требования к написанию и защите реферата: обозначена проблема и обоснована её актуальность, сделан краткий анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция, сформулированы выводы, тема раскрыта полностью, выдержан	отлично

	<p>объём, соблюдены требования к внешнему оформлению, даны правильные ответы на дополнительные вопросы. Реферат раскрывает поднятую проблематику в полном объеме.</p>	
	<p>Основные требования к реферату и его защите выполнены, но при этом допущены недочёты. В частности, имеются неточности в изложении материала; отсутствует логическая последовательность в суждениях; не выдержан объём реферата; имеются упущения в оформлении; на дополнительные вопросы при защите даны неполные ответы. В реферате имеются неточности и предметная область выступления раскрыта не в полной мере.</p>	хорошо
	<p>Имеются существенные отступления от требований к реферированию. В частности: тема освещена лишь частично; допущены фактические ошибки в содержании реферата или при ответе на дополнительные вопросы; во время защиты отсутствует вывод. В реферате не в полной степени раскрыт понятийный аппарат, имеются существенные неточности в процессе формирования выводов.</p>	удовлетворительно
	<p>Тема реферата не раскрыта, обнаруживается существенное непонимание проблемы. Тема реферата не раскрыта или выполнена не по существу ранее поставленного вопроса. Реферат не сдан / доклад не сделан.</p>	неудовлетворительно

### Критерии оценивания тестов

Средство контроля	Критерии оценивания – процент положительных ответов	Шкала оценивания
Тестирование	90-100	отлично
	70-89	хорошо
	40-69	удовлетворительно
	< 39	неудовлетворительно

### 3. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Средства оценивания в ходе промежуточной аттестации:

- вопросы для экзамена;
- тестовые задания к экзамену.

Коды компетенции	Результаты освоения ОПОП	Перечень планируемых результатов обучения по дисциплине

ОПК-5	Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.1. Знать: основы математики, физики, вычислительной техники и программирования. ОПК-1.2. Уметь: решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования. ОПК-1.3. Владеть: навыками теоретического и экспериментального исследования объектов профессиональной деятельности
-------	---	--

### 3.1. Вопросы к экзамену

1. Что такое интеллектуальный анализ данных (Data Mining)?
2. Чем Data Mining отличается от машинного обучения?
3. Какие основные этапы включает процесс KDD (Knowledge Discovery in Databases)?
4. Какие типы данных используются в интеллектуальном анализе данных?
5. Что такое предобработка данных и зачем она нужна?
6. Какие методы очистки данных существуют?
7. Что такое пропущенные значения и как с ними работать?
8. Какие способы нормализации данных вы знаете?
9. В чем суть масштабирования признаков?
10. Что такое выбросы и как их обнаружить?
11. Какие задачи решает интеллектуальный анализ данных?
12. Что такое классификация?
13. Что такое регрессия?
14. Что такое кластеризация?
15. Что такое ассоциативные правила?
16. В чем отличие обучения с учителем и без учителя?
17. Что такое обучающая и тестовая выборка?
18. Что такое переобучение модели?
19. Что такое недообучение модели?
20. Как бороться с переобучением?
21. Что такое алгоритм k-ближайших соседей?
22. Какие недостатки у метода kNN?
23. Что такое наивный байесовский классификатор?
24. На чем основан принцип Байеса?
25. Что такое дерево решений?
26. Какие критерии используются при построении дерева решений?
27. Что такое энтропия в контексте деревьев решений?
28. Что такое случайный лес?
29. В чем отличие случайного леса от одного дерева решений?
30. Что такое градиентный бустинг?
31. Что такое линейная регрессия?
32. Какие предположения лежат в основе линейной регрессии?
33. Что такое функция потерь?
34. Что такое метод наименьших квадратов?
35. Что такое логистическая регрессия?
36. Для чего используется сигмоидная функция?
37. Чем классификация отличается от регрессии?
38. Что такое ROC-кривая?

39. Что такое AUC?
40. Какие метрики качества моделей вы знаете?
41. Что такое кластеризация k-means?
42. Как работает алгоритм k-means?
43. Как выбирается число кластеров?
44. Что такое иерархическая кластеризация?
45. В чем разница между агломеративной и дивизивной кластеризацией?
46. Что такое DBSCAN?
47. Какие преимущества у DBSCAN?
48. Что такое плотность в кластеризации?
49. Какие проблемы возникают при кластеризации?
50. Что такое снижение размерности данных?
51. Что такое метод главных компонент (PCA)?
52. Зачем применяется PCA?
53. Что такое собственные значения и собственные векторы в PCA?
54. Что такое t-SNE?
55. В чем отличие PCA и t-SNE?
56. Что такое feature selection?
57. Какие методы отбора признаков существуют?
58. Что такое корреляция признаков?
59. Как влияет мультиколлинеарность на модели?
60. Какие реальные задачи решаются с помощью интеллектуального анализа данных?

### **3.2. Задания для зачета:**

#### **Задание 1**

Напишите модуль для нахождения НОД числа 8 и заданного 10-значного числа, не используя арифметические операции, но с использованием функций пользователя.

#### **Задание 2**

Создайте произвольный двухуровневый список из целых, вещественных и комплексных чисел. Выберите из него только комплексные числа и запишите их в кортеж.

#### **Задание 3**

Создайте множество из 100 первых элементов последовательности частичных сумм ряда Фибоначчи. Выведите число значащих цифр у элемента последовательности с заданным номером.

#### **Задание 4**

Создайте короткий скрипт, находящий все нули заданной алгебраической функции на заданном интервале (переменная формата float).

#### **Задание 5**

Скачайте любую статью (например, с [Navg.ru](http://Navg.ru)) и сформируйте из нее строку. Посчитайте частотности для всех слов, содержащихся в ней.

#### **Задание 6**

Создайте множество из случайных натуральных чисел размером в 10000 элементов. Осуществите из него случайную выборку 10 элементов без возвращения, а затем с возвращением, не используя методы `sample` и `choice`. Зафиксируйте время выполнения. Сделайте то же самое с использованием указанных методов модуля `random` и сравните скорости выполнения.

#### **Задание 7**

Создайте класс `Frac`, типичный экземпляр которого является обыкновенной дробью. Опишите методы обращения, сложения и умножения.

#### **Задание 8**

Постройте графики стандартного отклонения температур по месяцам года для Москвы и Анадыря, пользуясь данными [Gismeteo \(https://www.gismeteo.ru/diary\)](https://www.gismeteo.ru/diary).

#### **Задание 9**

Постройте три столбчатые диаграммы населения 5 стран мира (по вашему выбору) в 2000, 2010 и 2020 годах по данным ресурса worldometer.

#### **Задание 10**

Реализуйте функцию, которая принимает список чисел и возвращает медиану, не используя встроенные функции статистики.

#### **Задание 11**

Напишите программу, которая из заданной строки удаляет все знаки препинания и формирует список уникальных слов без использования библиотек NLP.

#### **Задание 12**

Создайте список из 1000 случайных чисел с нормальным распределением. Разделите его на три группы: ниже среднего, около среднего, выше среднего.

#### **Задание 13**

Дан вложенный список числовых данных (например, результаты измерений). Преобразуйте его в одномерный массив и удалите все выбросы по правилу  $3\sigma$ .

#### **Задание 14**

Реализуйте численное приближение производной функции на заданном интервале и постройте график её изменения.

#### **Задание 15**

Найдите все локальные экстремумы заданной функции на интервале методом перебора с шагом float.

#### **Задание 16**

Сгенерируйте 10 000 случайных значений и сравните выборки, полученные методами с возвращением и без возвращения. Оцените среднее и дисперсию для каждой выборки.\.

#### **Задание 17**

Напишите программу, моделирующую броски двух кубиков, и постройте распределение суммы значений.

#### **Задание 18**

Реализуйте простой класс Dataset, который хранит числовые данные и вычисляет среднее, дисперсию и стандартное отклонение.

#### **Задание 19**

Загрузите данные о температуре воздуха за год для двух городов. Постройте графики средней температуры по месяцам и сравните их.

#### **Задание 20**

Постройте гистограмму распределения доходов населения по выбранной стране и интерпретируйте форму распределения.

#### **Задание 21**

На основе открытых данных (например, World Bank) визуализируйте динамику населения 3 стран за 20 лет.

#### **Задание 22**

Реализуйте простейший алгоритм k-means для кластеризации двумерных данных.

#### **Задание 23**

Вычислите коэффициент корреляции Пирсона для двух наборов данных и сделайте вывод о зависимости.

#### **Задание 24**

Постройте линейную регрессию по заданным данным и оцените качество аппроксимации.

#### **Задание 25**

Дан список числовых значений с пропусками (None). Замените пропуски средним значением списка.

#### **Задание 26**

Удалите из списка все дубликаты, сохранив порядок элементов.

#### **Задание 27**

Нормализуйте список чисел в диапазон  $[0, 1]$  без использования сторонних библиотек.

**Задание 28**

Преобразуйте список строковых чисел в список float, обработав возможные ошибки преобразования.

**Задание 29**

Дан набор данных с шумом. Реализуйте простое сглаживание методом скользящего среднего.

**Задание 30**

Найдите моду, медиану и среднее значение для заданного набора данных без использования библиотек.

**Задание 31**

Постройте эмпирическое распределение случайной величины и сравните его с теоретическим нормальным распределением.

**Задание 32**

Вычислите дисперсию и стандартное отклонение вручную.

**Задание 33**

Определите асимметрию распределения (skewness) по формуле.

**Задание 34**

Сгенерируйте выборку и проверьте закон больших чисел на практике.

**Задание 35**

Посчитайте частоту биграмм в тексте.

**Задание 36**

Реализуйте простейший алгоритм удаления стоп-слов из текста.

**Задание 37**

Определите наиболее часто встречающиеся 10 слов в тексте.

**Задание 38**

Разделите текст на предложения и вычислите среднюю длину предложения.

**Задание 39**

Реализуйте простую метрику TF (term frequency) для слов.

**Задание 40**

Смоделируйте 10 000 подбрасываний монеты и оцените вероятность выпадения орла.

**Задание 41**

Смоделируйте случайный процесс случайного блуждания (random walk).

**Задание 42**

Оцените вероятность суммы двух кубиков  $\geq 8$  методом Монте-Карло.

**Задание 43**

Смоделируйте очередь обслуживания (простая дискретная модель).

**Задание 44**

Реализуйте генерацию случайных чисел с заданным распределением.

**Задание 45**

Постройте гистограмму распределения значений выборки.

**Задание 46**

Постройте boxplot (ящик с усами) и определите выбросы.

**Задание 47**

Постройте scatter plot для двух зависимых переменных.

**Задание 48**

Визуализируйте изменение значения функции во времени.

**Задание 49**

Постройте несколько графиков на одном полотне и сравните их поведение.

**Задание 50**

Вычислите корреляционную матрицу для набора данных из 3 переменных.

**Задание 51**

Определите наличие линейной зависимости между двумя наборами данных.

**Задание 52**

Разделите данные на обучающую и тестовую выборки вручную.

**Задание 53**

Реализуйте простую линейную регрессию через метод наименьших квадратов.

**Задание 54**

Оцените ошибку модели (MSE) для предсказанных значений.

**Задание 55**

Реализуйте k-means кластеризацию для двумерных данных.

**Задание 56**

Реализуйте алгоритм k-nearest neighbors (k-NN) для классификации.

**Задание 57**

Разбейте данные на кластеры и визуализируйте результат.

**Задание 58**

Определите оптимальное k методом перебора.

**Задание 59**

Сравните результаты кластеризации при разных начальных условиях.

**Задание 60**

Дана выборка числовых данных, описывающих поведение некоторого процесса (например, изменение показателя во времени). Разделите данные на интервалы и определите на каждом интервале среднее значение и дисперсию. Постройте график изменения этих характеристик и сделайте вывод о стабильности процесса.

#### 4. ОСНОВНЫЕ КРИТЕРИИ ОЦЕНИВАНИЯ ПРИ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

##### Шкала оценивания уровня сформированности компетенций (по пятибалльной системе) экзамен

Формируемые уровни освоения компетенций	Критерии оценивания	Шкала оценивания
Высокий уровень	Изложено правильное понимание вопроса, четко и самостоятельно дан исчерпывающий ответ, содержание раскрыто полно, профессионально, грамотно. Обучающимся усвоена взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии. Отражает успешное и систематическое применение навыков и умений по данной дисциплине в соответствии с ФГОС.	отлично

Базовый уровень	Изложено правильное понимание вопроса, дано достаточно подробное описание предмета ответа, приведены и раскрыты в тезисной форме основные понятия, относящиеся к предмету ответа. Ответ отражает полное знание учебно-программного материала, систематический характер знаний по дисциплине, а также наличие базового уровня овладения практическими умениями и навыками по данной дисциплине в соответствии с ФГОС	хорошо
Пороговый уровень	Ответ отражает теоретические знание основного учебно-программного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по профессии. Данная оценка может быть выставлена обучающемуся, допустившему неточности в ответе, но обладающими необходимыми знаниями для их устранения под руководством преподавателя, отмечен начальный уровень овладения практическими умениями и навыками по данной дисциплине в соответствии с ФГОС	удовлетворительно
Неудовлетворительный уровень	При ответе обучающегося обнаружено отсутствие знаний, умений и навыков и/или фрагментарные знания основного учебно-программного материала.	неудовлетворительно

**Текущий контроль и промежуточная аттестация** осуществляются в соответствии с «Положением о текущей и промежуточной аттестации обучающихся в Автономной некоммерческой организации «Образовательная организация высшего образования» «Университет экономики и управления».

Вид промежуточной аттестации – экзамен.

Форма проведения промежуточной аттестации – письменный экзамен.