

Спецификация конкурсных материалов

для проведения практического этапа Московского конкурса межпредметных навыков и знаний «Интеллектуальный мегаполис. Потенциал» в номинации «ИТ-класс» по направлению *Большие данные и технологии искусственного интеллекта*

1. Назначение конкурсных материалов

Материалы практического этапа Московского конкурса межпредметных навыков и знаний «Интеллектуальный мегаполис. Потенциал» (далее – Конкурс) предназначены для оценки уровня практической подготовки участников Конкурса.

2. Условия проведения

Практический этап Конкурса проводится в очной форме на базе ВУЗа. При выполнении работы обеспечивается строгое соблюдение порядка организации и проведения Конкурса. Используемое оборудование: компьютеры. На компьютерах установлена версия Python 3.9 или выше, а также интегрированные среды разработки (IDE) PyCharm и Visual Studio Code. Участники имеют доступ к необходимым библиотекам Python, включая *numpy*, *pandas* и *matplotlib*, которые настроены и готовы к использованию. Для работы с базами данных на компьютерах также установлена СУБД PostgreSQL и визуальный интерфейс pgAdmin или DBeaver.

3. Продолжительность выполнения

На выполнение заданий практического этапа Конкурса отводится **120** минут. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив ответственного от вуза. Мероприятие не продлевается на время отсутствия участника.

4. Содержание и структура

Индивидуальный вариант участника включает 7 заданий, базирующихся на содержании элективных курсов Введение в ИТ-специальность и Информационные технологии.

5. Система оценивания

Задание считается выполненным, если ответ участника совпал с эталоном. Максимальный балл за выполнение всех заданий – 60 баллов.

6. Приложения

1. План конкурсных материалов для проведения практического этапа Конкурса.
2. Демонстрационный вариант конкурсных заданий практического этапа Конкурса.
3. Критерии оценки заданий

**План конкурсных материалов
для проведения практического этапа Конкурса**

№ задания	Уровень сложности	Уникальные кодификаторы Конкурса	Контролируемые требования к проверяемым умениям	Балл
1	<i>Базовый</i>	Математические основы технологий искусственного интеллекта	Знание и понимание основных понятий математики, комбинаторики, теории вероятности необходимых для применения технологий искусственного интеллекта.	6
2	<i>Базовый</i>	График функции Гистограммы Распределения	Умение визуализировать данные с помощью графиков и гистограмм, интерпретировать полученные результаты. Знание основных типов графиков и их применения для анализа распределения данных. Понимание распределений данных (например, нормальное распределение, биномиальное распределение и т.п.)	8
3	<i>Базовый</i>	Введение в Python Базовые операции	Основы программирования на Python, включая работу с переменными, типами данных, операциями и базовыми структурами данных (списки, словари и т.д.).	6

4	<i>Базовый</i>	Введение в SQL Примеры в PostgreSQL	Умение формировать и оптимизировать SQL-запросы, знание основ работы с базами данных. Практическое применение SQL в PostgreSQL.	6
5	<i>Повышенный</i>	Библиотека numpy Библиотека pandas Библиотека matplotlib	Умение использовать библиотеку NumPy для создания и преобразования массивов, выполнения арифметических операций и генерации случайных чисел. Умение применять Pandas для работы с табличными данными: создание и модификация DataFrame, фильтрация, группировка, обработка пропущенных значений и объединение данных. Умение использовать Matplotlib для визуализации данных: построение графиков, настройка осей и оформления.	10
6	<i>Повышенный</i>	Введение в машинное обучение	Понимание основных концепций машинного обучения, таких как регрессия, классификация, методы обучения с учителем и без учителя. Знание алгоритмов, таких как линейная регрессия, деревья решений.	9

7	<i>Повышенный</i>	<p>Нормализация данных</p> <p>Проектирование базы данных в PostgreSQL</p>	<p>Знание принципов нормализации данных (1NF, 2NF, 3NF), проектирование базы данных с учётом этих принципов. Умение разрабатывать структуры баз данных в PostgreSQL, оптимизировать запросы и поддерживать целостность данных.</p>	15
Сумма баллов:				60

**Демонстрационный вариант конкурсных заданий
практического этапа Конкурса**

Пример состава задания практического этапа Конкурса.

Ответы к задачам, записанные в файлы, необходимо сохранять в папку

Потенциал_ИТ_ФамилияИО

Задача 1.

В магазине 40% товаров - электроника, 30% - одежда, 30% - продукты.

5% электроники, 10% одежды и 1% продуктов имеют дефекты.

Какова вероятность выбрать дефектный товар?

В ответ запишите развернутое решение. Для обоснования воспользуйтесь символами: $P(A)$ для обозначения вероятности.

Запишите свое решение в файл: задача_1.txt

Задача 2.

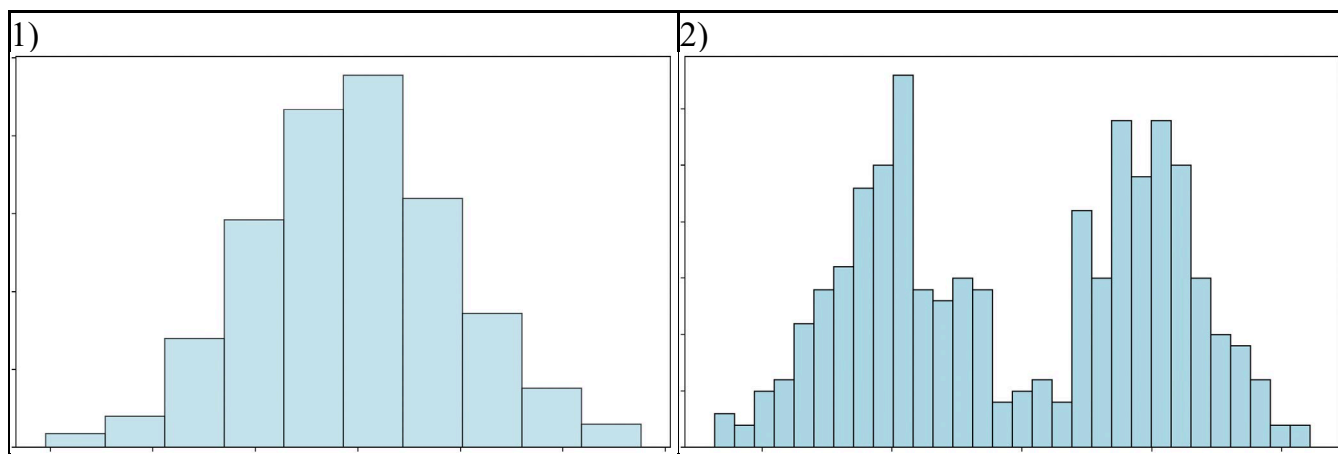
Перед вами четыре гистограммы, каждая из которых отображает распределение данных из одной из четырех выборок, описанных ниже. На гистограммах:

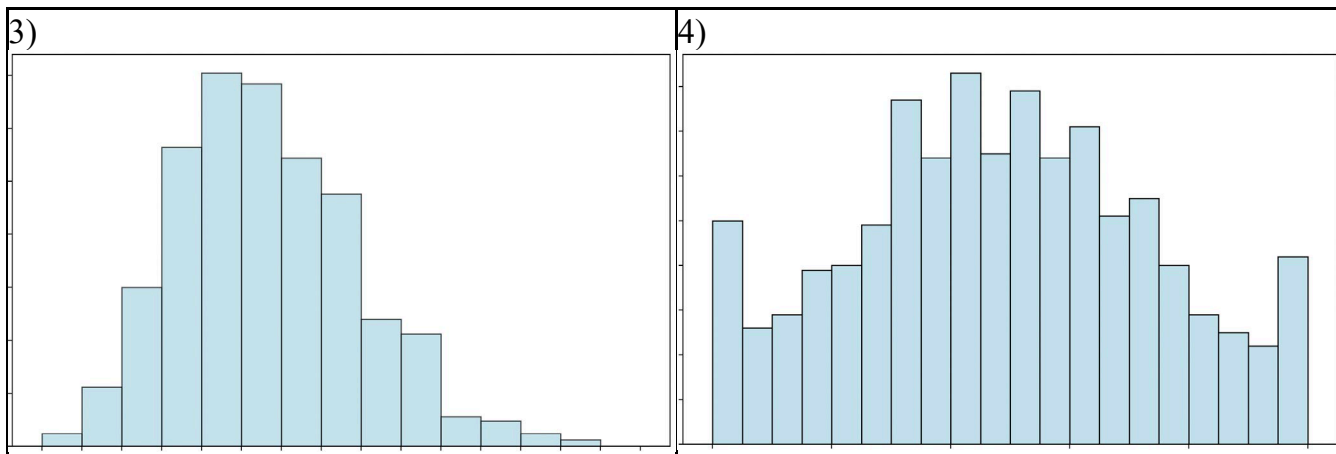
- **Ось абсцисс** показывает интервалы значений.
- **Ось ординат** отражает частоту.

Ваша задача — сравнить эти гистограммы с описаниями выборок и определить, какая из них соответствует каждому графику.

Свой ответ обоснуйте.

Запишите свое решение в файл: задача_2.txt





А) Выборка оценок студентов по 100-бальной системе, полученных на экзамене по физике в течение одного учебного года. Данная выборка включает в себя небольшое количество студентов.

Б) Выборка количества очков, набранных спортсменами в серии из десяти игр в школьной лиге. Максимальное количество очков, набранных одним спортсменом, составляет 15.

В) Выборка показателей веса учеников старших классов в общеобразовательной школе, измеренных в начале учебного года.

Г) Выборка результатов опроса о предпочтениях в еде среди учеников в школе. Участникам было предложено выбрать из 10 категорий продуктов.

Задача 3.

Ввод: стандартный ввод

Вывод: стандартный вывод

Напишите программу, которая считает количество вхождений первого символа в введенную строку. (Гарантируется, что в строке будет хоть один символ)

Формат вывода: В строке <введенная строка> найдено <количество (тип int)> символов.

Запишите свое решение в файл: *task_3.py*

Задача 4.

В базе данных есть таблица **students**, содержащая следующие поля: **id**, **name**, **age**, **grade**. Напишите SQL-запрос, который выберет имена всех студентов, которым 16 лет.

Запишите свое решение в файл: *задача_4.txt*

Задача 5.

Вам предоставлен набор данных (*library_books.csv*), содержащий информацию о книгах в библиотеке. Этот набор данных включает следующие столбцы:

- Название книги
- Автор
- Год издания
- Жанр
- Количество выданных за последний год
- Общее количество экземпляров

Используя библиотеку pandas, проведите анализ данных и предоставьте отчет по следующим критериям:

1. Определите топ-3 самых популярных книг (по количеству выданных за последний год).
2. Постройте один график, показывающий среднее количество выданных экземпляров для каждого жанра.
3. Убедитесь, что график имеет название и подписанные оси.

Формат вывода:

1)

Топ-3 популярных книг:

1. "Название книги 1" (Автор 1) - количество выданных экземпляров: X
2. "Название книги 2" (Автор 2) - количество выданных экземпляров: Y
3. ...

Решение задачи и полученные файлы сохраните в папке **задача_5**.

Содержимое папки оформить следующим образом:

Пункт задачи	Описание файла	Кодировка файла	Название файла
1	В файл необходимо поместить вывод программы в указанном формате	UTF-8	задача_5_1.txt
2	График, полученный в ходе решения задачи		задача_5_2.png (задача_5_2.jpg)
Общий файл с программой	Полный код программы с комментариями	UTF-8	task_5.py

Задача 6.

В этом коде используется алгоритм К-ближайших соседей для классификации данных из датасета Iris. Однако в процессе предсказания допущена ошибка: найдите и исправьте ее. В качестве ответа запишите исправленный код. Аргументируйте свой ответ.

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

# Загружаем данные
iris = load_iris()
X, y = iris.data, iris.target

# Разделяем данные
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Обучаем модель
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train, y_train)

# Предсказания
y_pred = knn.predict(X_train)
print("Предсказания:", y_pred)
```

Код (решение) запишете в файл task_6.py или задача_6.txt.

Задача 7.

Вы работаете в компании, занимающейся продажей книг. Вам необходимо спроектировать базу данных для хранения информации о книгах, авторах и заказах. Ваша задача — создать схему базы данных, которая будет удобной для хранения и извлечения данных.

Условия задачи:

Книги:

- Каждая книга имеет уникальный идентификатор, название, год издания и цену.
- У каждой книги есть один автор.

Авторы:

- Каждый автор имеет уникальный идентификатор, имя и фамилию.
- Один автор может написать несколько книг.

Заказы:

- Каждый заказ имеет уникальный идентификатор, дату заказа и сумму.
- Один заказ может включать несколько книг, но каждая книга в заказе может быть только одна.

Задания:

1. Нормализация данных:

- Определите таблицы, которые необходимо создать для хранения информации о книгах, авторах и заказах (опишите текстом структуру БД).
- Опишите, как вы будете устранять избыточность данных.

2. Создание структуры базы данных:

- Напишите SQL-код для создания таблиц в PostgreSQL с использованием подходящих типов данных, первичных и внешних ключей.
- Обеспечьте поддержку целостности данных с помощью ограничений.

3. Запросы:

- Напишите SQL-запрос для получения списка всех книг с указанием их авторов.

Примеры возможного заполнения таблиц:

Таблица Авторы:

AuthorID	FirstName	LastName
1	George	Orwell
2	J.K.	Rowling
3	F. Scott	Fitzgerald

Таблица Книги:

BookID	Title	YearPublished	Price	AuthorID
1	1984	1949	9.99	1
2	Animal Farm	1945	7.99	1
3	Harry Potter and the Sorcerer's Stone	1997	14.99	2
4	Harry Potter and the Chamber of Secrets	1998	15.99	2
5	The Great Gatsby	1925	10.99	3

Таблица Заказы:

OrderID	OrderDate	TotalAmount
1	2024-08-01	25.98
2	2024-08-03	15.99
3	2024-08-05	7.99

Решение задачи и полученные файлы сохраните в папке задача_7
Файлы в папке оформите следующим образом:

Пункт задачи	Что должно быть в файле	Кодировка файла	Название файла
Нормализация данных	В файле опишите структуру БД	UTF-8	задача_7_1.txt
Создание структуры БД	В файл добавьте SQL запрос для создания таблиц	UTF-8	задача_7_2.txt
Запросы	Добавьте SQL-запрос	UTF-8	задача_7_3.txt

Критерии оценки заданий

Задача 1.

Баллы	Критерии оценки
0	Дан неверный ответ, не предоставлено никакого объяснения.
2	Дан неправильный ответ, но представлены некоторые шаги решения.
4	Дан правильный ответ, но объяснение недостаточно ясно.
6	Дан правильный ответ и представлено полное и логичное решение.

Задача 2.

Баллы	Критерии оценки
0	Не сопоставлен ни один график с описанием
2	Правильно сопоставлен 1 график, но не связаны остальные
4	Правильно сопоставлены 2 графика
6	Правильно сопоставлено 3 графика
8	Правильно сопоставлены все графики

Задача 3.

Баллы	Критерии оценки
0	Код отсутствует или содержит критические ошибки, мешающие его выполнению.
	Не представлены правильные решения задачи.
	Отсутствие попытки анализа задачи.
3	Код работает, но содержит незначительные ошибки или недочеты.
	Решение задачи выполнено, но требует доработки.
	Основные операции в коде присутствуют, но неправильно использованы.
	Часть логики реализована, но не завершена.
6	Код работает без ошибок и оптимально решает задачу.
	Решение задачи логично и четко структурировано.
	Использованы правильные конструкции и подходы, код легко читаем.
	В случае необходимости, предоставлены комментарии для улучшения понимания.

Задача 4.

Баллы	Критерии оценки
0	Дан неверный ответ, не предоставлено работающего SQL-запроса.
3	Дан правильный SQL-запрос, но он может содержать незначительные ошибки или быть не полностью оптимизирован.
	Запрос выполняет задачу, но требует улучшений в плане производительности или правильности вывода.
6	Дан полностью правильный и оптимизированный SQL-запрос.
	Запрос выполняет задачу корректно и эффективно, без ошибок и с оптимальной производительностью.

Задача 5.

Критерий	Описание	Баллы
Анализ топ-3 самых популярных книг	Определены и правильно указаны топ-3 книги по количеству выданных за последний год.	4
График	Построен график, показывающий среднее количество выданных экземпляров для каждого жанра. График должен быть четким и информативным.	3
Чистота и структурированность кода	Код написан аккуратно, с комментариями и правильной структурой, что облегчает понимание и повторное использование.	3
Всего баллов		10

Задача 6.

Баллы	Критерии оценки
0	Код отсутствует или содержит критические ошибки, мешающие его выполнению.
	Не представлены правильные решения задачи.
	Отсутствие попытки анализа ошибки или объяснений к исправлениям.
5	Код исправлен, но остаются незначительные ошибки или недочеты.
	Исправление кода представлено, но отсутствует обоснование выбора решения.
9	Код написан правильно и оптимально, решает задачу без ошибок.
	Ответы полные и содержат детальное объяснение всех исправлений, демонстрируя глубокое понимание алгоритма.

Задача 7.

Пункт задачи	Критерии оценки	Баллы
1. Нормализация данных		6
1.1 Определение таблиц и атрибутов	Определены все необходимые таблицы и атрибуты, соблюдены принципы 3NF	3
1.2 Устранение избыточности и зависимостей	Корректно устранены избыточности, правильно установлены связи и ключи	3
2. Создание структуры базы данных		7
2.1 Создание таблиц с правильными типами данных	Использованы корректные типы данных для всех атрибутов	3
2.2 Создание первичных и внешних ключей	Корректно определены первичные и внешние ключи	2
2.3 Обеспечение целостности данных	Применены ограничения (NOT NULL, UNIQUE, CHECK и т.д.)	2
3. Примеры запросов		2
3.1 Список книг с указанием авторов	Запрос корректно возвращает все книги с их авторами	2
Всего баллов за задачу		15