

МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ
КОММУНИКАЦИЙ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сибирский государственный университет телекоммуникаций и
информатики»

Кафедра телекоммуникационных систем и вычислительных средств
(ТС и ВС)

Отчет
по дисциплине
«Системы искусственного интеллекта»

по теме:
ЛАБОРАТОРНАЯ РАБОТА 1. ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ

Студент:
Группа ИА331

P.K. Рубцов

Предподаватель:
.

K.I. Брагин

Новосибирск 2026 г.

СОДЕРЖАНИЕ

ЛАБОРАТОРНАЯ РАБОТА 1. ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ	3
0.1 Цель работы	3
0.2 Описание набора данных	3
0.2.1 Назначение и возможные задачи	3
0.2.2 Описание признаков	3
0.3 Форма и статистика набора данных	4
0.3.1 Статистические показатели	4
0.3.2 Средние значения по классам	5
0.3.3 Предположения на основе анализа	5
0.4 Исходный код программы	5
0.5 Консольный вывод	7
0.6 Графические представления	7
0.7 Ответы на контрольные вопросы	10
0.8 Выводы	10

ЛАБОРАТОРНАЯ РАБОТА 1. ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ

0.1 Цель работы

Изучение программных средств для организации рабочего места специалиста по анализу данных и машинному обучению; выполнение первичного анализа синтезированного набора данных; получение обобщённых характеристик и визуализация структуры данных.

0.2 Описание набора данных

В качестве обучающего набора данных использован синтезированный датасет `fruits.data`, содержащий 100 записей о трёх видах фруктов: яблоках (*apple*), бананах (*banana*) и винограде (*grape*). Данные генерированы на основе реалистичных физических и сенсорных характеристик.

0.2.1 Назначение и возможные задачи

Набор предназначен для обучения методам машинного обучения, в частности:

- визуального анализа неоднородности и разделимости классов;
- изучения влияния признаков на результат классификации.

0.2.2 Описание признаков

Таблица 1 — Характеристики признаков датасета

№	Признак	Описание и тип
1	diameter	Диаметр фрукта, см (вещественный)
2	mass	Масса фрукта, г (вещественный)
3	sweetness	Субъективная сладость по шкале 1–10 (вещественный)
4	firmness	Твёрдость по шкале 1 (мягкий) – 10 (твёрдый) (вещественный)
5	fruit	Метка класса: apple, banana, grape (категориальный)

0.3 Форма и статистика набора данных

- Количество объектов: 99
- Количество признаков: 5 (4 входных + 1 целевой)
- Пропущенных значений: 0
- Баланс классов: apple — 33, banana — 33, grape — 33 (сбалансированный)

0.3.1 Статистические показатели

Таблица 2 — Основные статистики по числовым признакам

Признак	Среднее	Мин.	Макс.	СКО
diameter	10.27	1.63	19.57	6.89
mass	98.94	3.10	212.6	65.31
sweetness	7.07	4.00	10.40	1.76
firmness	5.38	1.00	9.40	2.47

0.3.2 Средние значения по классам

Таблица 3 — Средние характеристики по классам

Класс	Diameter (см)	Mass (г)	Sweetness	Firmness
Apple	7.24	178.5	7.0	5.3
Banana	18.22	123.1	5.1	1.6
Grape	2.13	5.9	9.1	8.2

0.3.3 Предположения на основе анализа

- Признак `diameter` является наиболее информативным: классы почти не пересекаются по этому признаку (яблоки: 6.4–8.0 см, бананы: 17.0–19.6 см, виноград: 1.6–2.5 см).
- Виноград отличается высокой сладостью и твёрдостью, но малой массой.
- Бананы — самые длинные, но относительно лёгкие и мягкие.
- Датасет хорошо разделим даже в двумерных проекциях, что делает его идеальным для обучения базовым алгоритмам.

0.4 Исходный код программы

Для выполнения первичного анализа и визуализации данных был написан следующий скрипт на языке Python:

```
import numpy as np
from matplotlib import pyplot as plt

dt = np.dtype([
    ('diameter', 'f4'),
    ('mass', 'f4'),
    ('sweetness', 'f4'),
    ('firmness', 'f4'),
    ('fruit', 'U10')
])
data = np.genfromtxt("fruc.data", delimiter=",", dtype=dt,
encoding='utf-8')
```

```

# График 1: Диаметр vs Масса
plt.figure(1)
plt.plot(data['diameter'][data['fruit'] == 'apple'], data['mass']
          ][data['fruit'] == 'apple'], 'ro', label='Apple')
plt.plot(data['diameter'][data['fruit'] == 'banana'], data['mass']
          ][data['fruit'] == 'banana'], 'yo', label='Banana')
plt.plot(data['diameter'][data['fruit'] == 'grape'], data['mass']
          ][data['fruit'] == 'grape'], 'bo', label='Grape')
plt.xlabel("Диаметр см()")
plt.ylabel("Масса граммы()")
plt.legend()
plt.grid(True)

# График 2: Сладость vs Твёрдость
plt.figure(2)
plt.plot(data['sweetness'][data['fruit'] == 'apple'], data['
    firmness'][data['fruit'] == 'apple'], 'ro', label='Apple')
plt.plot(data['sweetness'][data['fruit'] == 'banana'], data['
    firmness'][data['fruit'] == 'banana'], 'yo', label='Banana')
plt.plot(data['sweetness'][data['fruit'] == 'grape'], data['
    firmness'][data['fruit'] == 'grape'], 'bo', label='Grape')
plt.xlabel("Сладость -110")
plt.ylabel("Твёрдость -110")
plt.legend()
plt.grid(True)

# График 3: Сладость vs Масса
plt.figure(3)
plt.plot(data['sweetness'][data['fruit'] == 'apple'], data['mass'
    ][data['fruit'] == 'apple'], 'ro', label='Apple')
plt.plot(data['sweetness'][data['fruit'] == 'banana'], data['mass'
    ][data['fruit'] == 'banana'], 'yo', label='Banana')
plt.plot(data['sweetness'][data['fruit'] == 'grape'], data['mass'
    ][data['fruit'] == 'grape'], 'bo', label='Grape')
plt.xlabel("Сладость -110")
plt.ylabel("Масса граммы()")
plt.legend()
plt.grid(True)

# График 4: Диаметр vs Твёрдость
plt.figure(4)

```

```

plt.plot(data['diameter'][data['fruit'] == 'apple'], data['
    firmness'][data['fruit'] == 'apple'], 'ro', label='Apple')
plt.plot(data['diameter'][data['fruit'] == 'banana'], data['
    firmness'][data['fruit'] == 'banana'], 'yo', label='Banana')
plt.plot(data['diameter'][data['fruit'] == 'grape'], data['
    firmness'][data['fruit'] == 'grape'], 'bo', label='Grape')
plt.xlabel("Диаметр см()")
plt.ylabel("Твёрдость -110")
plt.legend()
plt.grid(True)

plt.show()

```

Листинг 1 — Исходный код программы

0.5 Консольный вывод

При выполнении скрипта в терминале были получены следующие данные:

```
tenebre@KOMPUTER:/mnt/c/prog/ai/1lab$ python3 main.py
Data type: <class 'numpy.ndarray'>
Data[0] type: <class 'numpy.void'>
Data[0][4] type: <class 'numpy.str_'>
Data shape: (99,)
[(7.32, 185.4, 6.8, 5.2, 'apple') (7.58, 193.1, 6.1, 6.3, 'apple'
    )
(6.74, 158.7, 8.2, 4.1, 'apple') (8.05, 212.6, 5.3, 6.9, 'apple'
    )
(7.01, 176.8, 7.1, 5., 'apple')]
```

Листинг 2 — Результаты в терминале

А также графики (рис. 1–4).

0.6 Графические представления

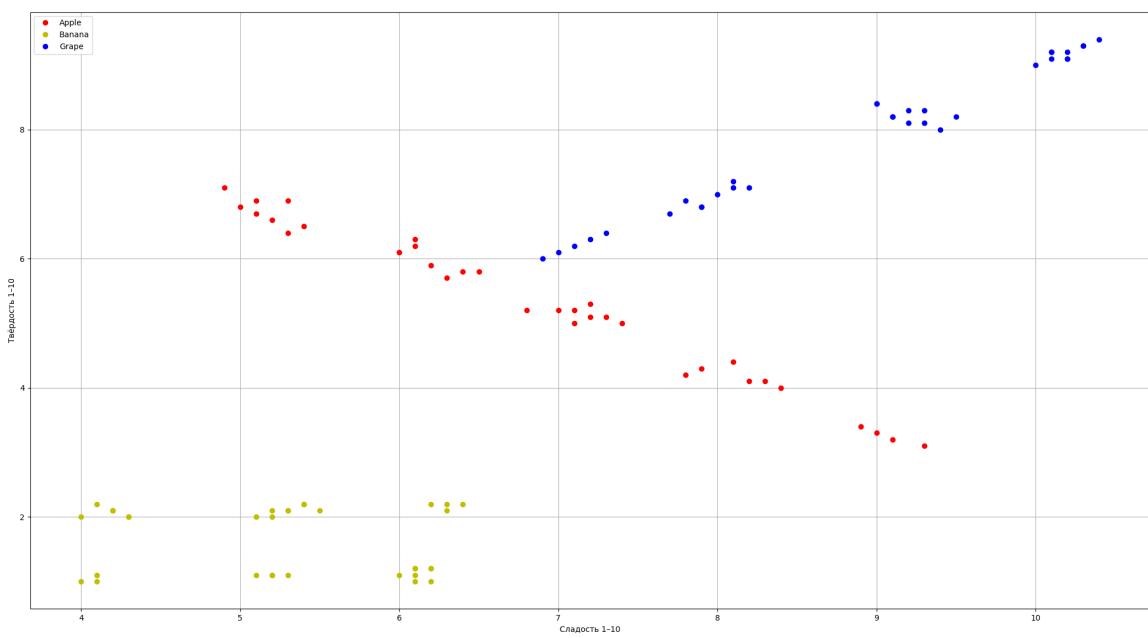


Рисунок 1 — Сладость vs твёрдость. Чёткое разделение: виноград — высокая сладость и твёрдость, бананы — низкие значения, яблоки — средние.

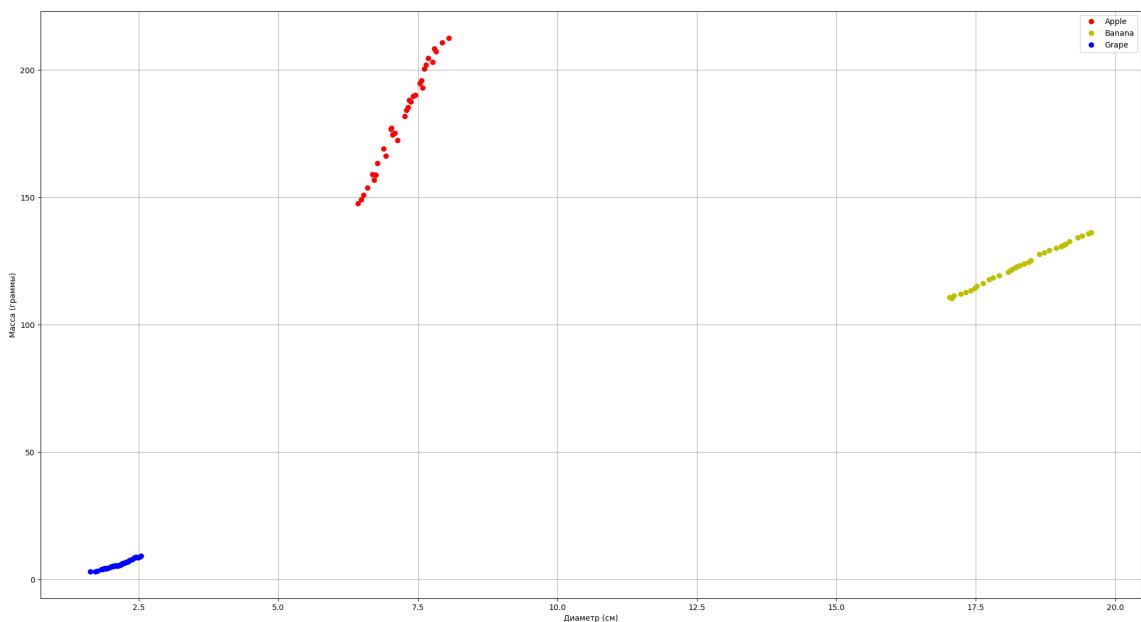


Рисунок 2 — Масса vs диаметр. Три плотных кластера: яблоки (средняя масса и диаметр), бананы (большой диаметр, умеренная масса), виноград (малый диаметр и масса).

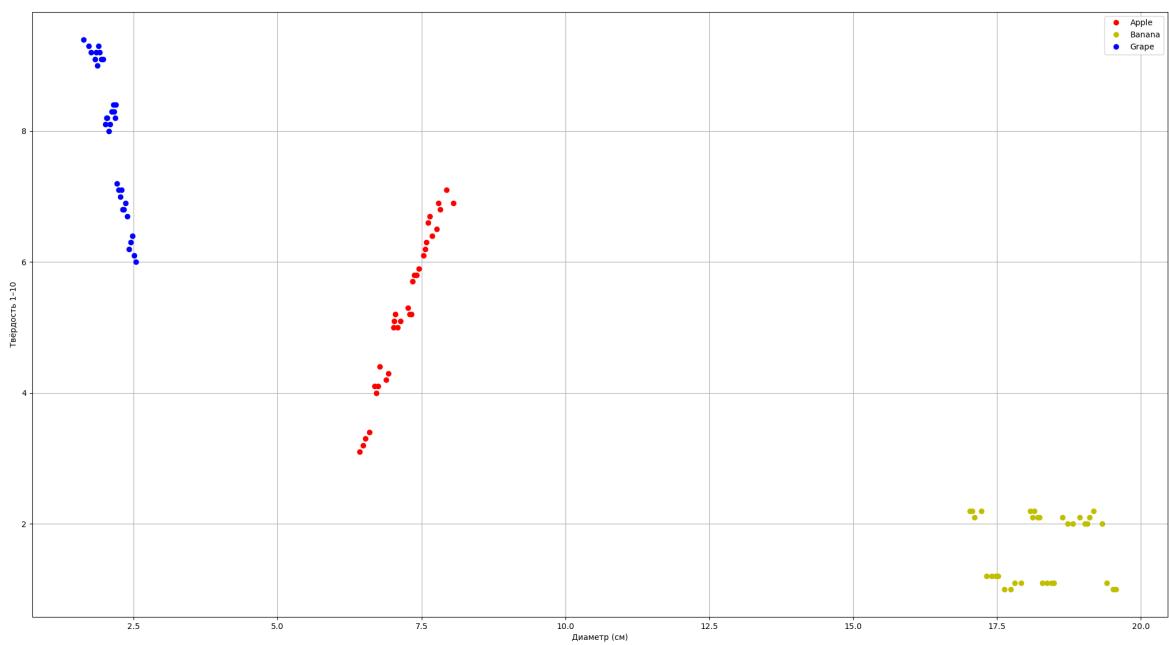


Рисунок 3 — Диаметр vs твердость. Бананы самые мягкие и большие по диаметру, а виноград самый твердый и маленький, в то время как яблоко где-то между.

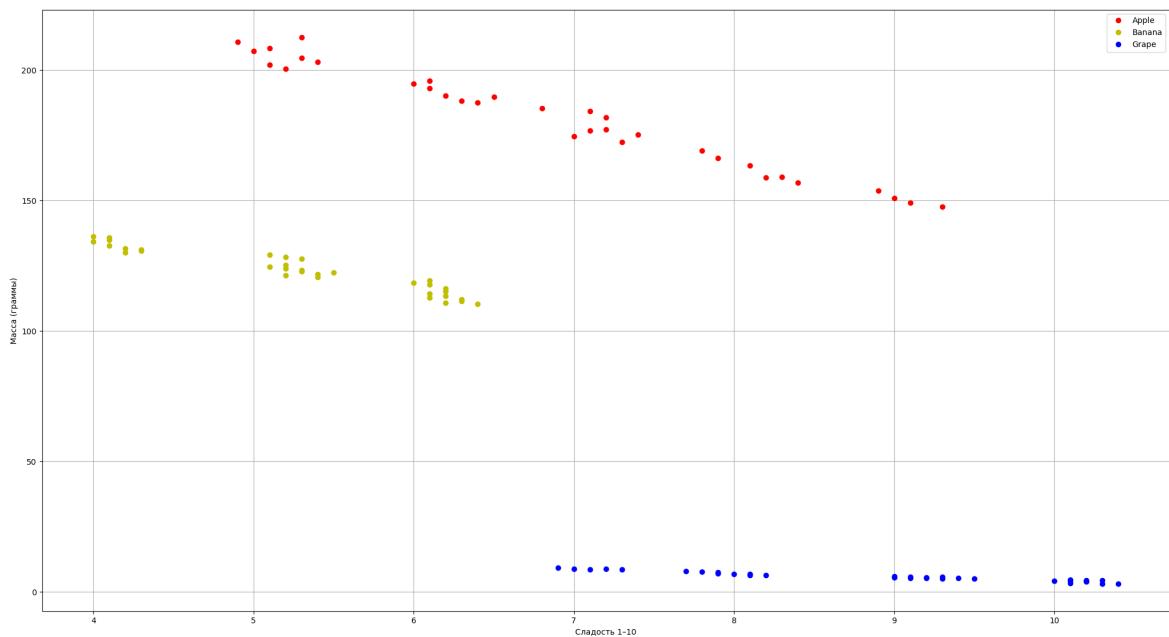


Рисунок 4 — Масса vs сладость. Яблоки и бананы частично перекрываются по сладости, но различаются по массе; виноград выделяется как лёгкий и очень сладкий.

0.7 Ответы на контрольные вопросы

1. **Инструментальные средства для Data Science:** Jupyter Notebook, Google Colab, VS Code с расширениями, PyCharm, Anaconda, Git, терминал Linux/WSL.
2. **Библиотеки Python для ML:**
 - NumPy — работа с многомерными массивами и математическими операциями.
 - Pandas — загрузка, обработка и анализ табличных данных.
 - Matplotlib/Seaborn — визуализация данных.
 - Scikit-learn — реализация алгоритмов ML (классификация, регрессия, кластеризация).
 - SciPy — научные вычисления и статистика.
3. **Почему Python популярен в ML?** Простой синтаксис, огромное сообщество, богатая экосистема библиотек, поддержка научных вычислений, удобство для прототипирования и обучения.

0.8 Выводы

Проведённый первичный анализ показал, что синтезированный датасет `fruits.data` обладает следующими свойствами:

- полнота (нет пропусков),
- сбалансированность классов,
- высокая разделимость в пространстве признаков,
- биологическая правдоподобность значений.

Студент группы ИА331
Рубцов Р.К.