

Министерство образования Республики Беларусь
Учреждение образования
«Брестский государственный технический университет»
ФАКУЛЬТЕТ ЭЛЕКТРОННО-ИНФОРМАЦИОННЫХ СИСТЕМ
Кафедра интеллектуальных информационных технологий

Отчет по лабораторной работе №1

Выполнил
А.В. Горобец,
студент группы АС66
Проверил
А. А. Крощенко,
ст. преп. кафедры ИИТ,
« __ » _____ 2025 г.

Брест 2025

Цель работы: Получить практические навыки работы с данными с использованием библиотек Pandas для манипуляции и Matplotlib для визуализации.

Научиться выполнять основные шаги предварительной обработки данных, такие как очистка, нормализация и работа с различными типами признаков.

Вариант 3

Задание 1. Загрузите данные и проверьте, есть ли в них пропущенные значения.

```
import pandas as pd
df = pd.read_csv(r'C:\Users\Anton\Downloads\iris.csv')
pd.set_option('display.max_rows', None) # Показывать все строки
pd.set_option('display.max_columns', None) # Показывать все столбцы
pd.set_option('display.width', None) # Без ограничения по ширине
pd.set_option('display.max_colwidth', None) # Полная ширина столбцов
print(df)
print("\nПроверка на пропущенные значения:")
print(df.isnull().sum())
```

```
Проверка на пропущенные значения:
sepal.length    0
sepal.width     0
petal.length    0
petal.width     0
variety         0
dtype: int64
```

Задание 2. Выведите количество образцов каждого вида ириса.

```
import pandas as pd
df = pd.read_csv(r'C:\Users\Anton\Downloads\iris.csv')
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
print("\nКоличество образцов каждого вида ириса:")
print(df['variety'].value_counts())
```

```
Количество образцов каждого вида ириса:
variety
Setosa      50
Versicolor  50
Virginica   50
Name: count, dtype: int64
```

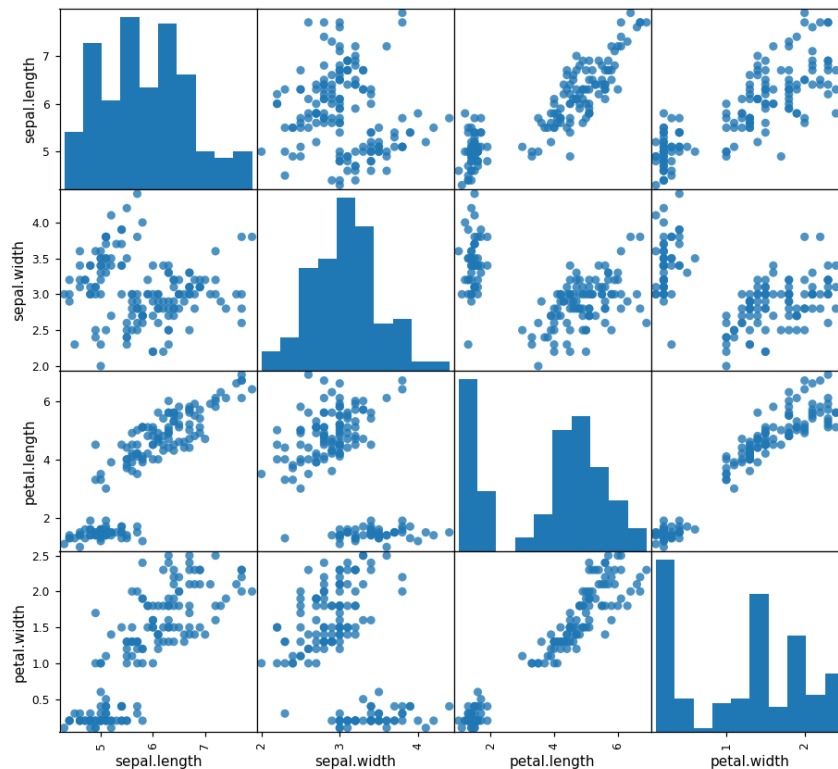
Задание 3. Постройте парные диаграммы рассеяния (pair plot) для всех признаков, чтобы визуально оценить их разделимость.

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')

plt.scatter(df['age'], df['thalach'], c=df['target'], cmap='coolwarm')

plt.xlabel('Возраст')
plt.ylabel('Максимальный пульс')
plt.show(block=True)
```

Парные диаграммы рассеяния для признаков Iris



Задание 4. Для каждого вида ириса рассчитайте среднее значение по каждому из четырех признаков.

```
import pandas as pd
df = pd.read_csv(r'C:\Users\Anton\Downloads\iris.csv')
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
mean_by_variety = df.groupby('variety').mean(numeric_only=True)
print("\nСредние значения признаков по каждому виду ириса:")
print(mean_by_variety)
```

```
Средние значения признаков по каждому виду ириса:
      sepal.length  sepal.width  petal.length  petal.width
variety
Setosa           5.006         3.428         1.462         0.246
Versicolor       5.936         2.770         4.260         1.326
Virginica        6.588         2.974         5.552         2.026
```

Задание 5. Создайте "ящик с усами" (box plot) для признака Petal Length (cm), чтобы сравнить его распределение по разным видам ирисов.

```
import pandas as pd
import matplotlib.pyplot as plt

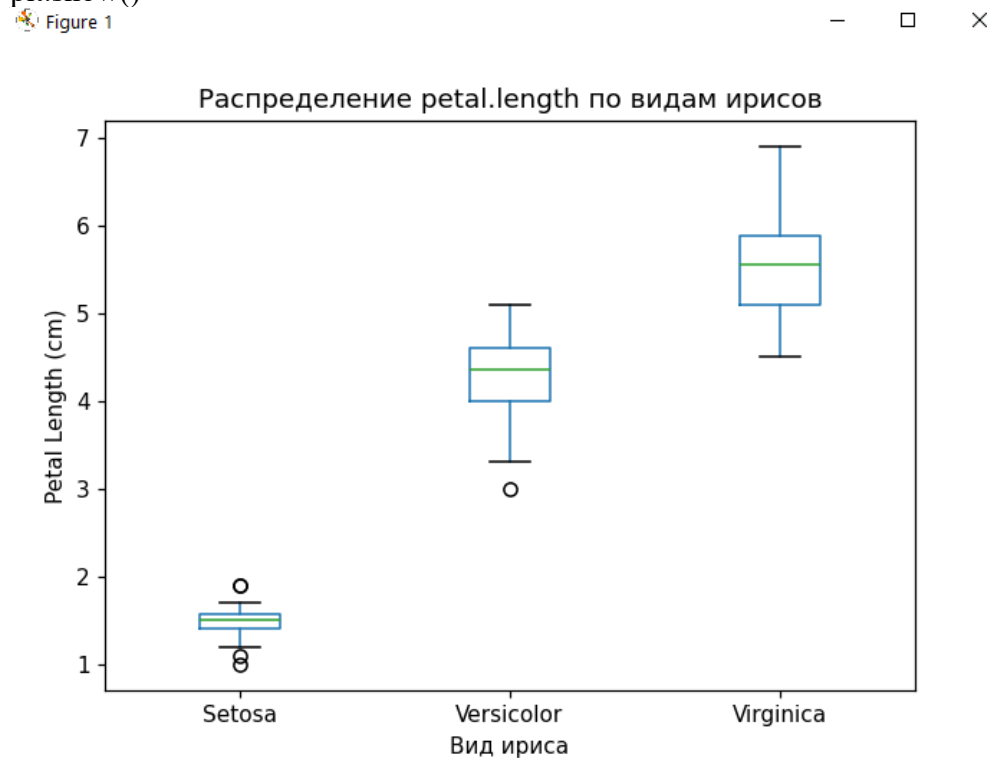
# Загрузка данных
df = pd.read_csv(r'C:\Users\Anton\Downloads\iris.csv')

# Настройки отображения
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
```

```
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)

# Построение box plot для признака petal.length
df.boxplot(column='petal.length', by='variety', grid=False)
```

```
plt.title('Распределение petal.length по видам ирисов')
plt.suptitle("") # Убираем автоматический заголовок
plt.xlabel('Вид ириса')
plt.ylabel('Petal Length (cm)')
plt.show()
```



Задание 6. Стандартизируйте данные (приведите к нулевому среднему и единичному стандартному отклонению).

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')

features_to_normalize = ['age', 'trestbps', 'chol', 'thalach']

scaler = MinMaxScaler()

df[features_to_normalize] = scaler.fit_transform(df[features_to_normalize])

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)

print(df)
```

143	1.159173	0.328414	1.217458	1.448832	Virginica
144	1.038005	0.558611	1.103783	1.712096	Virginica
145	1.038005	-0.131979	0.819596	1.448832	Virginica
146	0.553333	-1.282963	0.705921	0.922303	Virginica
147	0.795669	-0.131979	0.819596	1.053935	Virginica
148	0.432165	0.788808	0.933271	1.448832	Virginica
149	0.068662	-0.131979	0.762758	0.790671	Virginica

Вывод: я получил практические навыки работы с данными с использованием библиотек Pandas для манипуляции и Matplotlib для визуализации.

Научился выполнять основные шаги предварительной обработки данных, такие как очистка, нормализация и работа с различными типами признаков.