МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Э. Баумана

Факультет «Информатика и системы управления»
Кафедра «Систем обработки информации и управления»

# ОТЧЕТ

**Лабораторная работа №___1____**
по дисциплине «Методы машинного обучения»

Тема: «Создание "истории о данных"»

ИСПОЛНИТЕЛЬ: Ли Яцзинь
ФИО
группа ИУ5-25
_____
подпись
"___"_____2024 г.

ПРЕПОДАВАТЕЛЬ: _____Гапанюк_Ю .Е___
ФИО
_____
подпись
"___"_____2024 г.

Москва - 2024

**Задание**

Выбрать набор данных (датасет).

. Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:

1. История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.
2. На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.
3. Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.
4. Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.

. История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.

**Сформировать отчет и разместить его в своем репозитории на github.**

## Импорт библиотек

```
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(style="white",color_codes=True)
```

## Описание набора данных

Набор данных Iris использовался в R.A. Классическую статью Фишера 1936 года «Использование множественных измерений в таксономических задачах» также можно найти в репозитории машинного обучения UCI.

Он включает в себя три вида ирисов по 50 образцов каждый, а также некоторые свойства каждого цветка. Один вид цветка линейно отделим от двух других, но два других не отделимы линейно друг от друга.

Столбцы в этом наборе данных:

- Id
- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species
-

## Загрузка датасета

```
data = pd.read_csv("/content/Iris.csv")
print(data.head())
```

```
   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      Species
0   1            5.1           3.5            1.4           0.2  Iris-setosa
1   2            4.9           3.0            1.4           0.2  Iris-setosa
2   3            4.7           3.2            1.3           0.2  Iris-setosa
3   4            4.6           3.1            1.5           0.2  Iris-setosa
4   5            5.0           3.6            1.4           0.2  Iris-setosa
```

## Шаг1.Гистограмма （Histogram）

```
data_without_id = data.drop('Id', axis=1)    # 删除名为 'ID' 的列
data_without_id.plot(kind='hist', subplots=True, layout=(2, 2), figsize=(10, 10))
```

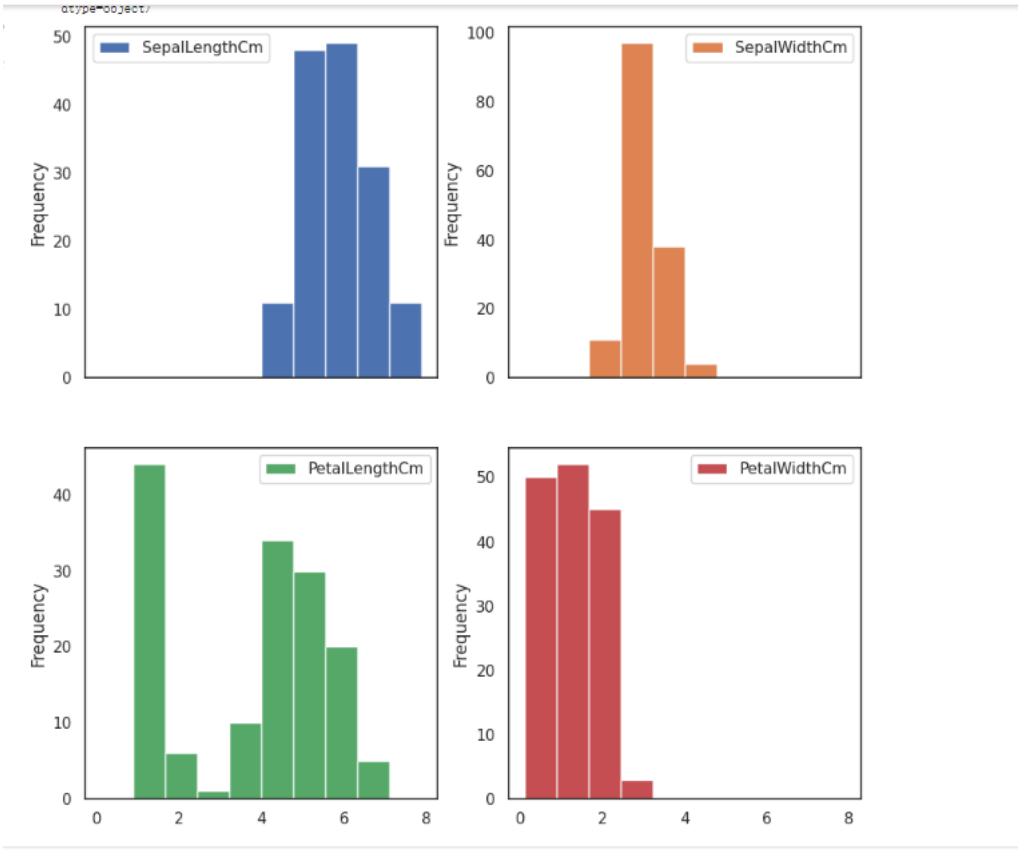


Рис- 1: Гистограмма четырех атрибутов цветка ириса

## Шаг2.График рассеяния(Scatter plot)

нарисовать диаграмму рассеяния двух переменных (длина и ширина чашелистика) при построении диаграммы рассеяния.

```
#设置绘图的种类,scatter表示散点图，x轴使用萼片的长度，y轴使用萼片的宽度
data.plot(kind="scatter",x="SepalLengthCm",y="SepalWidthCm")
plt.show()
```

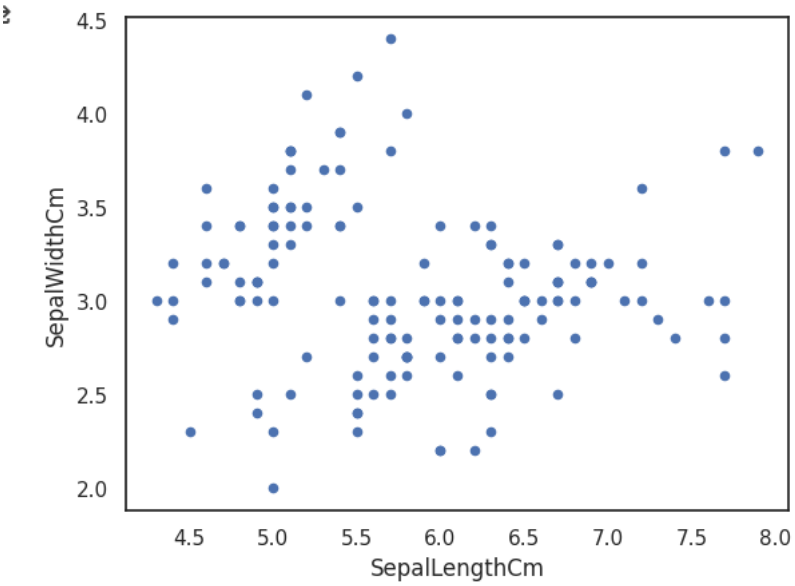


Рис- 2: неудачная график рассеяния

Используйте разные цвета, чтобы обозначить разные виды цветов ириса.

```
face = sns.FacetGrid(data,hue="Species",height=5)
face.map(plt.scatter,"SepalLengthCm","SepalWidthCm")
face.add_legend()
plt.show()
```

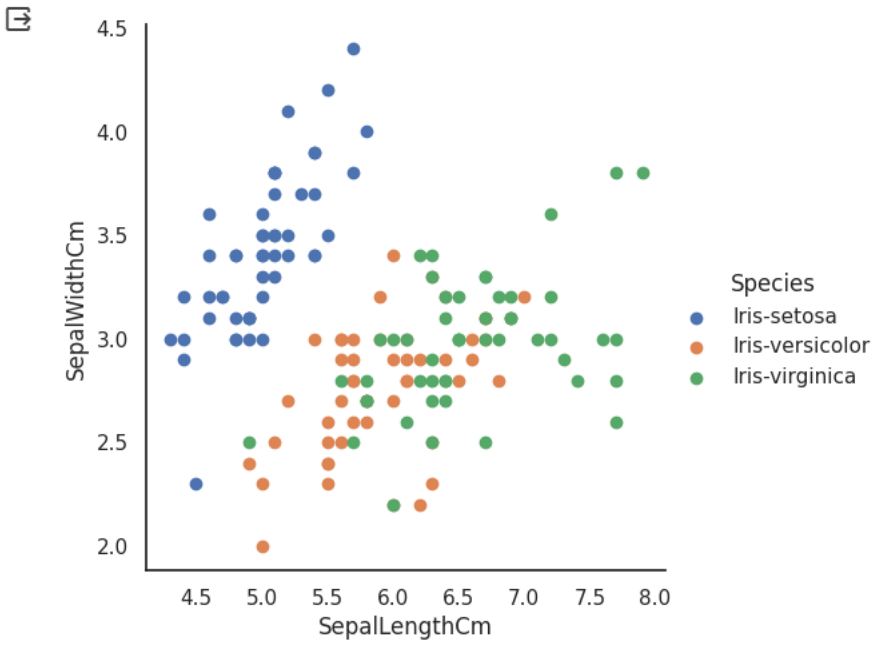


Рис- 3: удачная график рассеяния

**Шаг3.коробочный сюжет(boxplot)**

Нарисуйте коробчатую диаграмму в зависимости от длины цветков ириса.

```
#  箱型图
sns.boxplot(x="Species",y="SepalLengthCm",data=data)
plt.show()
```

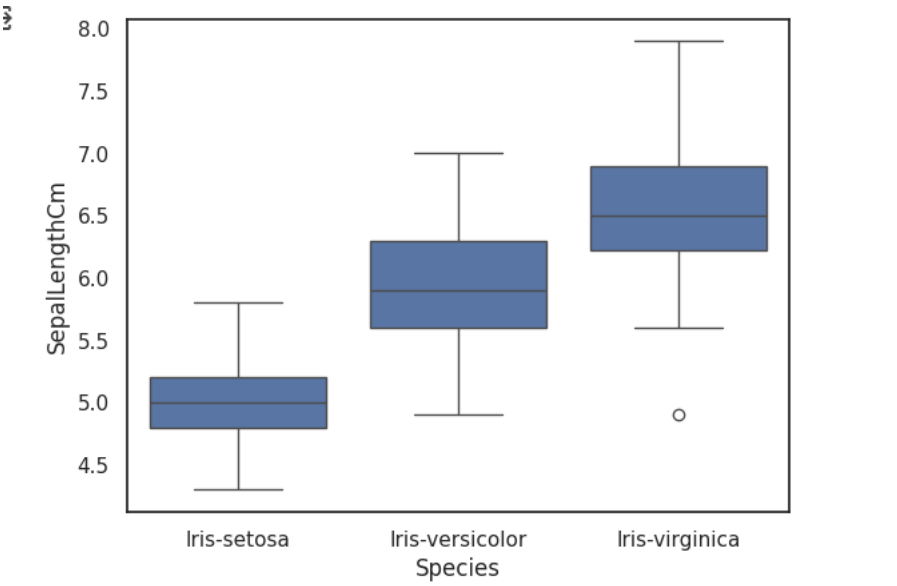


Рис- 4: неудачная коробочный сюжет

График на основе коробчатого графика:

```
ax  =  sns.boxplot(x="Species",y="SepalLengthCm",data=data)
#在箱形图的基础上进行描点，设置jitter为True保证点不会落在同一条直线上
ax  =  sns.stripplot(x="Species",y="SepalLengthCm",data=data,jitter=True,edgecolor="gray")
plt.show()
```

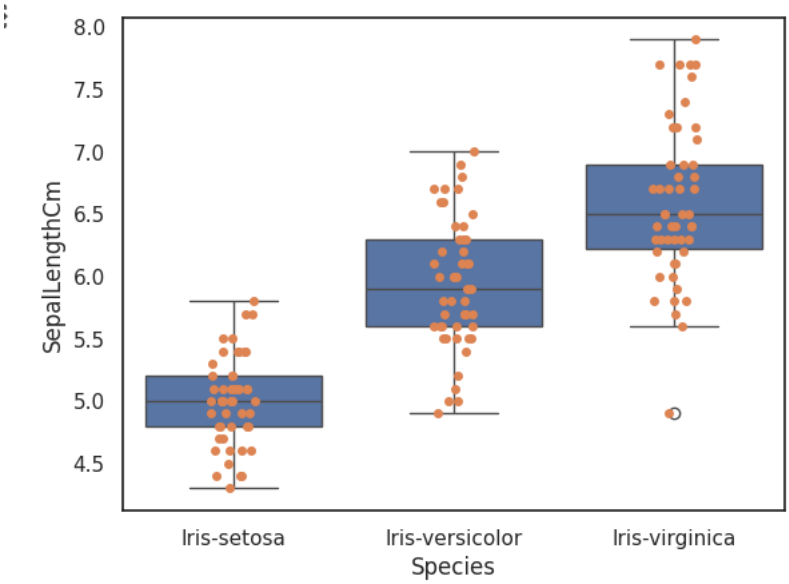


Рис- 5: удачная коробочный сюжет

## Шаг 4.Нарисуйте схему скрипки（Violin diagram）

```
sns.violinplot(x="Species",y="SepalLengthCm",data=data,gridsize=6)
plt.show()
```

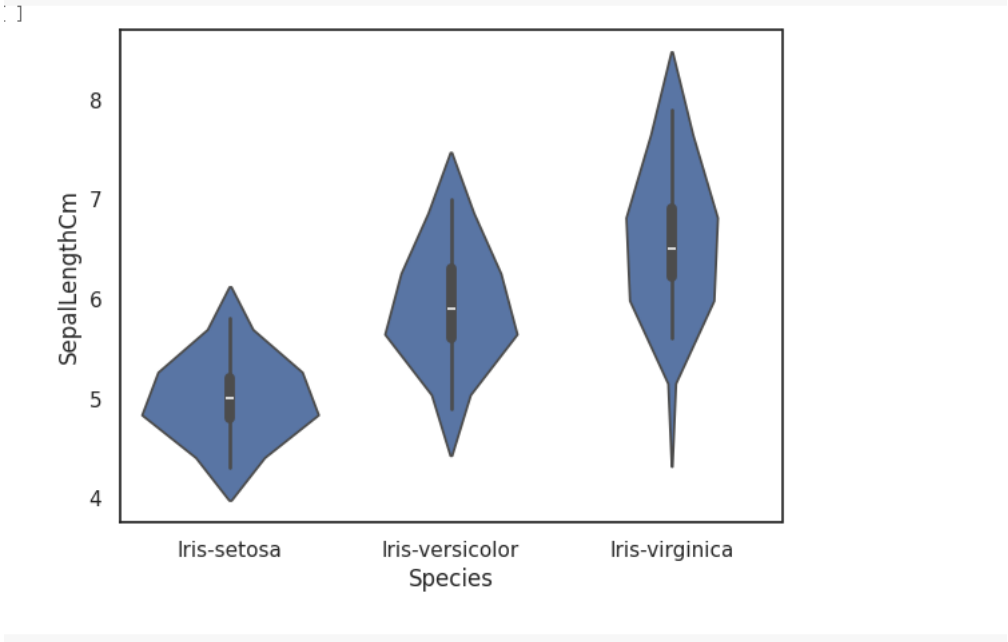


Рис- 6: Схема скрипки

## Шаг 5：Оценка плотности ядра（Plot Kernel Density Estimate）

```
face = sns.FacetGrid(data,hue="Species",height=6)
face.map(sns.kdeplot,"SepalLengthCm")
face.add_legend()
plt.show()
```

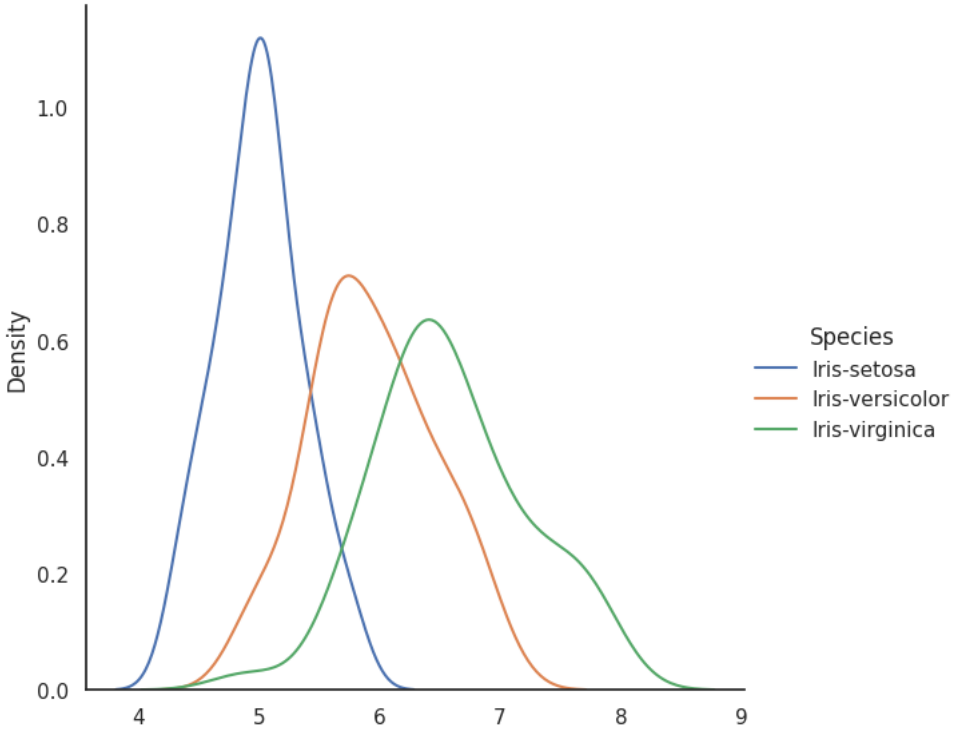


Рис- 7：График оценки плотности ядра