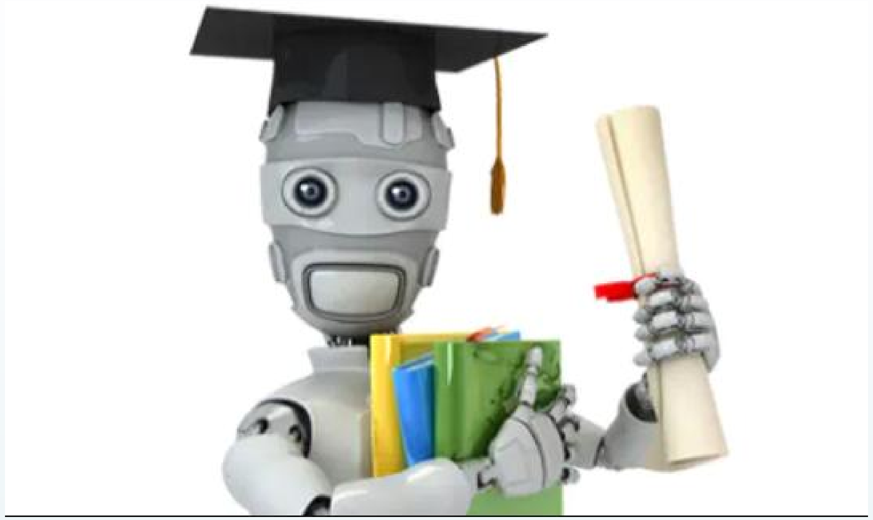
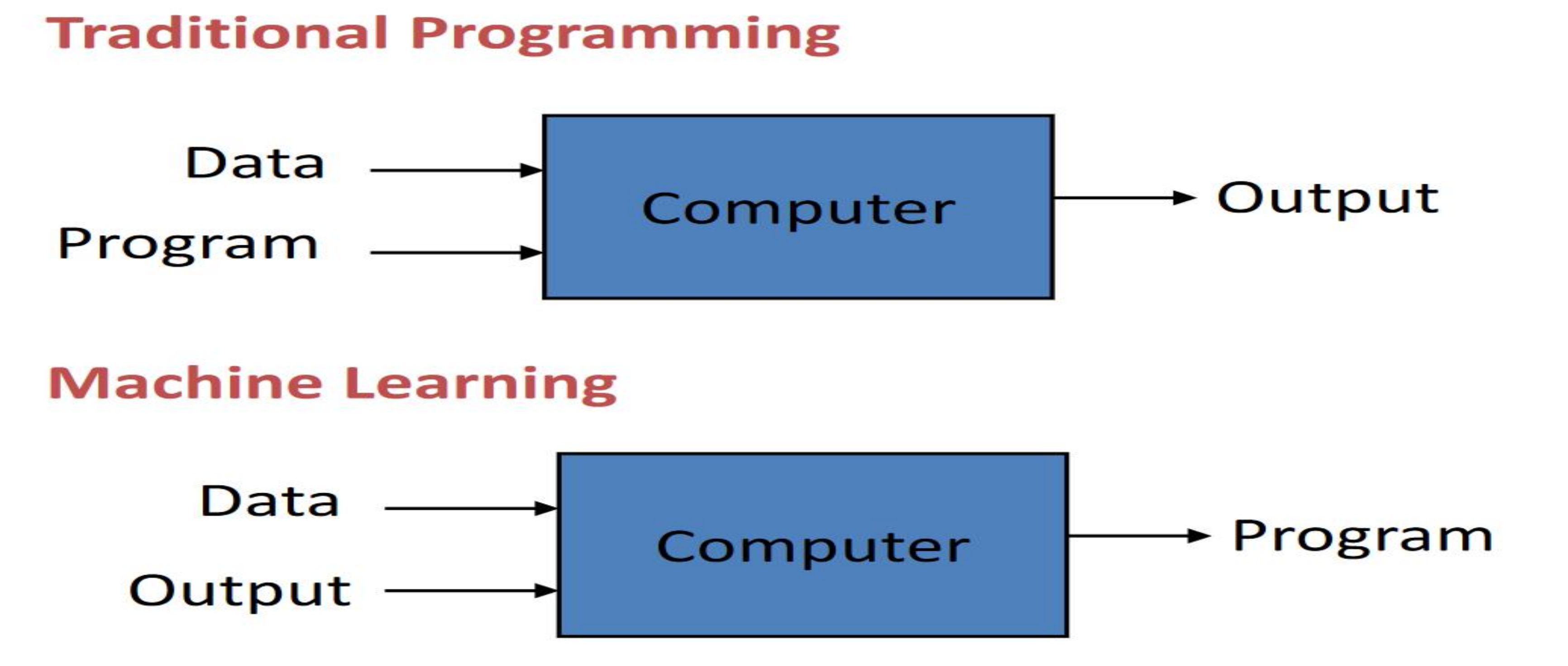


# Introduction To Machine Learning

# What is Machine Learning?

"Learning is any process by which a system improves performance from experience."
-Herbert Simon



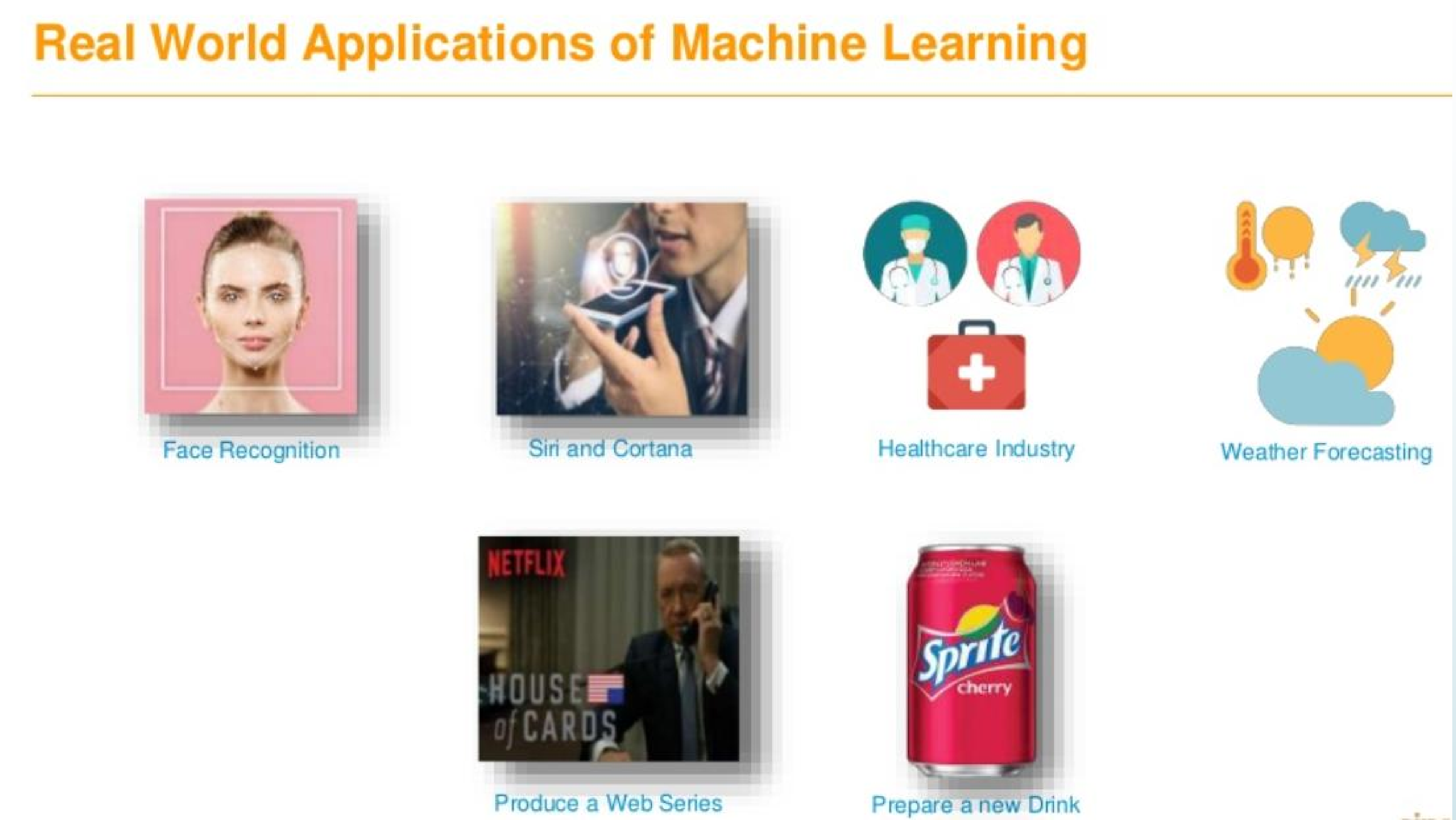
# When Do We Use Machine Learning?

ML is usedwhen:

- Human expertise doesnot exist(navigating on Mars)

- Humans can’t explain their expertise
( speechrecognition)

- Models are based on huge amounts of data



# Real World Applications of Machine Learning

Face Recognition

Siri and Cortana

Healthcare Industry

Weather Forecasting

Produce a Web Series

Prepare a new Drink

# Types of Learning:

- Supervised (inductive) learning–Given: training data + desired outputs (labels).

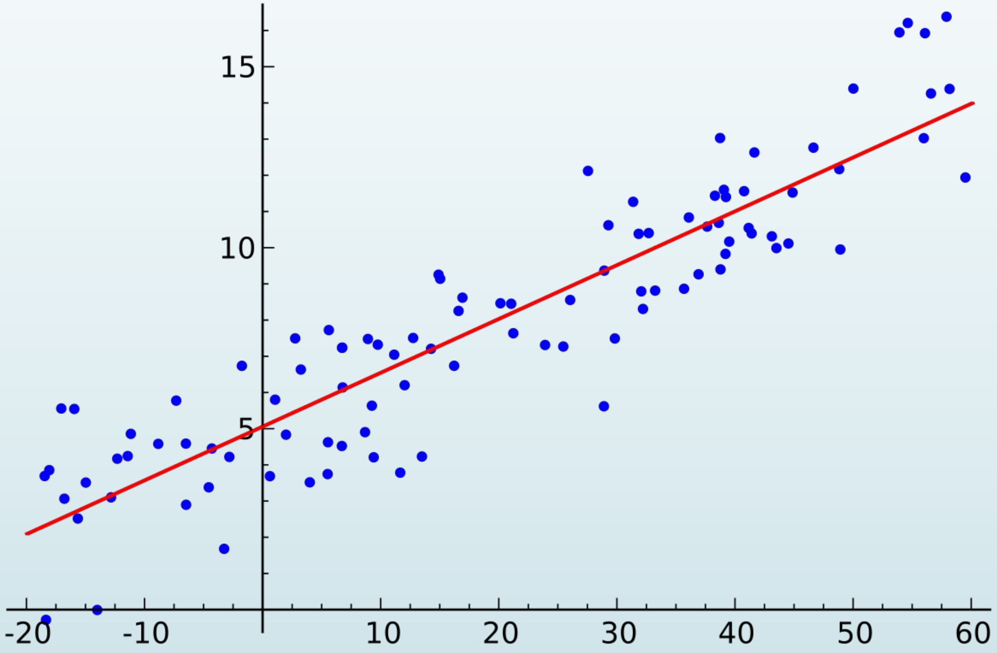
- Unsupervised learning–Given: training data (without desired outputs).

- Reinforcement learning–Rewards from sequence of actions.

# Supervised Learning:

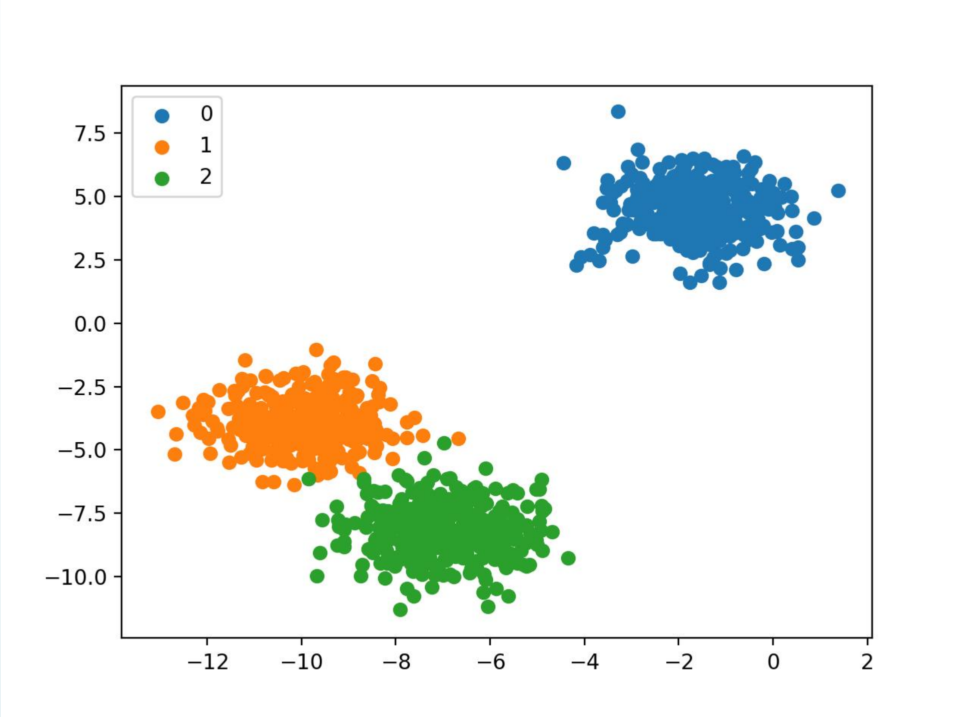
1)Regression

- Given (x1, y1), (x2, y2), ..., (xn, yn)
- Learn a function f(x) to predict y given x–yaxis real-valued == regression



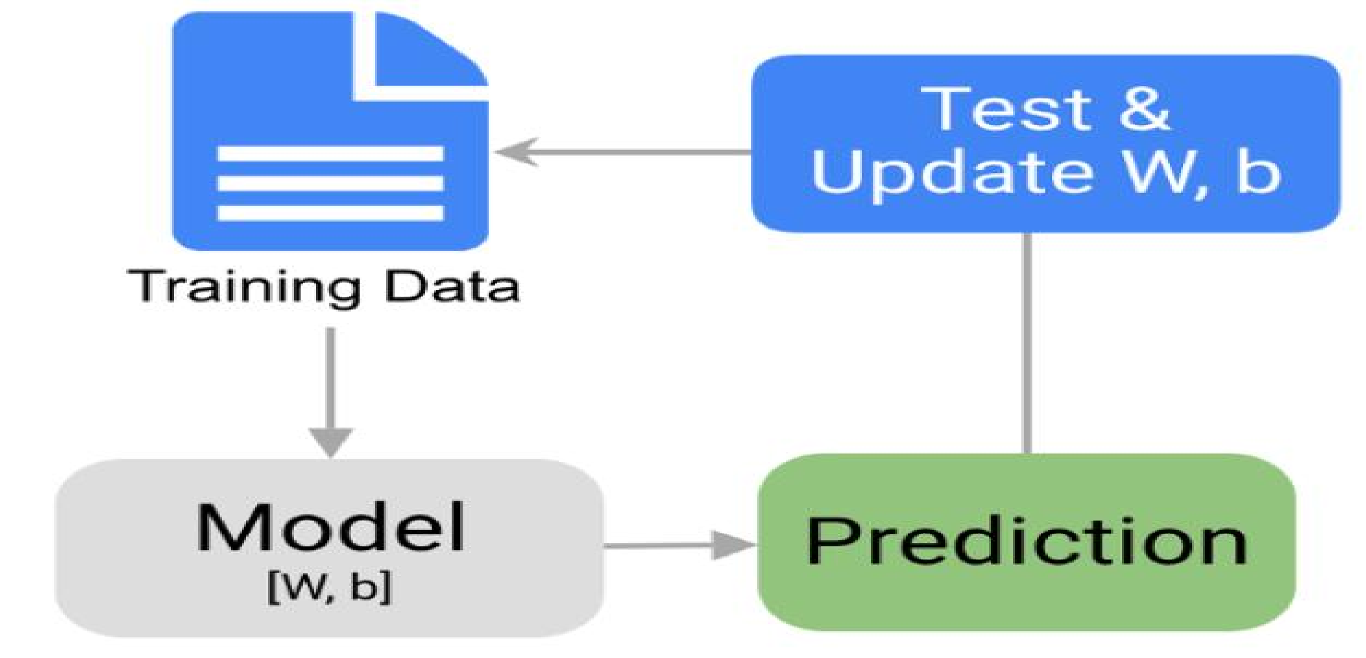
## 2)Classification

- Given (x1, y1), (x2, y2), ..., (xn, yn)
- Learn a function f(x) to predict y given x–yaxis categorical == classification



## Steps to build model:

1 - Data Collection
2 - Data Preparation
3 - Choose a Model
4 - Train the Model
5 - Evaluate the Model
6 - Make Predictions



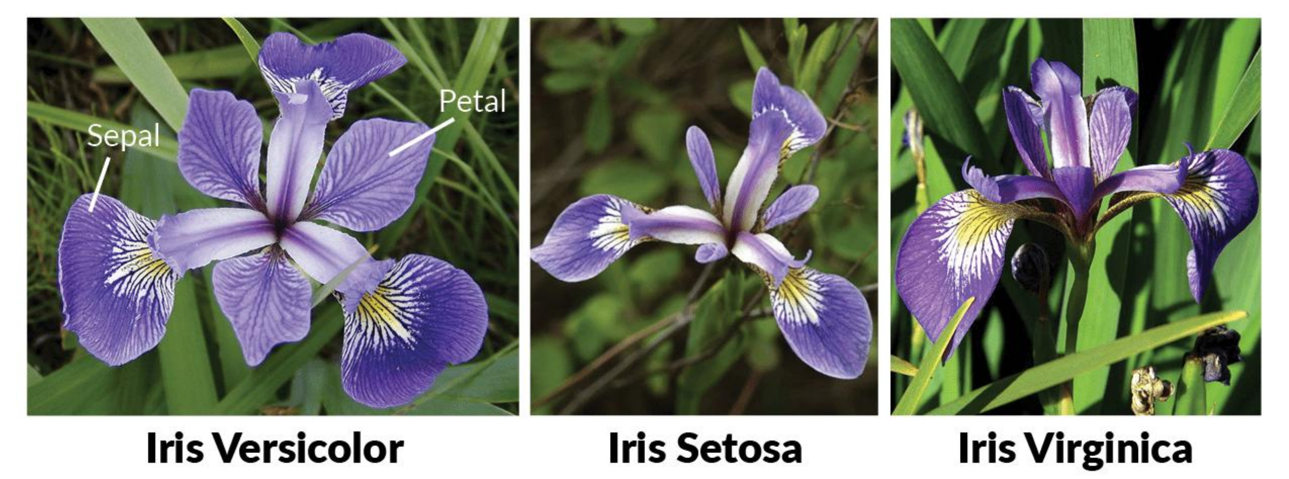
# Python libraries used in Machine Learning:

1)Numpy
2)Pandas
3)Scikit-learn
4)Matplotlib
5)TensorFlow
6)Keras
7)PyTorch
8)Scipy
9)Seaborn

# Iris Flower Data Set

- The *Iris* flower data set or Fisher's *Iris* data set is a multivariate data set introduced by Ronald Fisher in his 1936.

  It is sometimes called Anderson's *Iris* data set because Edgar Anderson collected the data to quantify the morphologic variation of *Iris* flowers of three related species. The use of this data set in cluster analysis is however uncommon, since the data set only contains two clusters with rather obvious separation.



- One of the clusters contains *Iris setosa*, while the other cluster contains both *Iris virginica* and *Iris versicolor* and is not separable without the species information Fisher used.

  It is multivariate(more than 2 dependent variable) data set Study of three related Iris flowers species. Data set contain 50 sample of each species(Iris-Setosa, Iris-Virginica, IrisVersicolor)

Features Used :

1. Sepal length in cm
2. Sepal width in cm
3. Petal length in cm
4. Petal width in cm

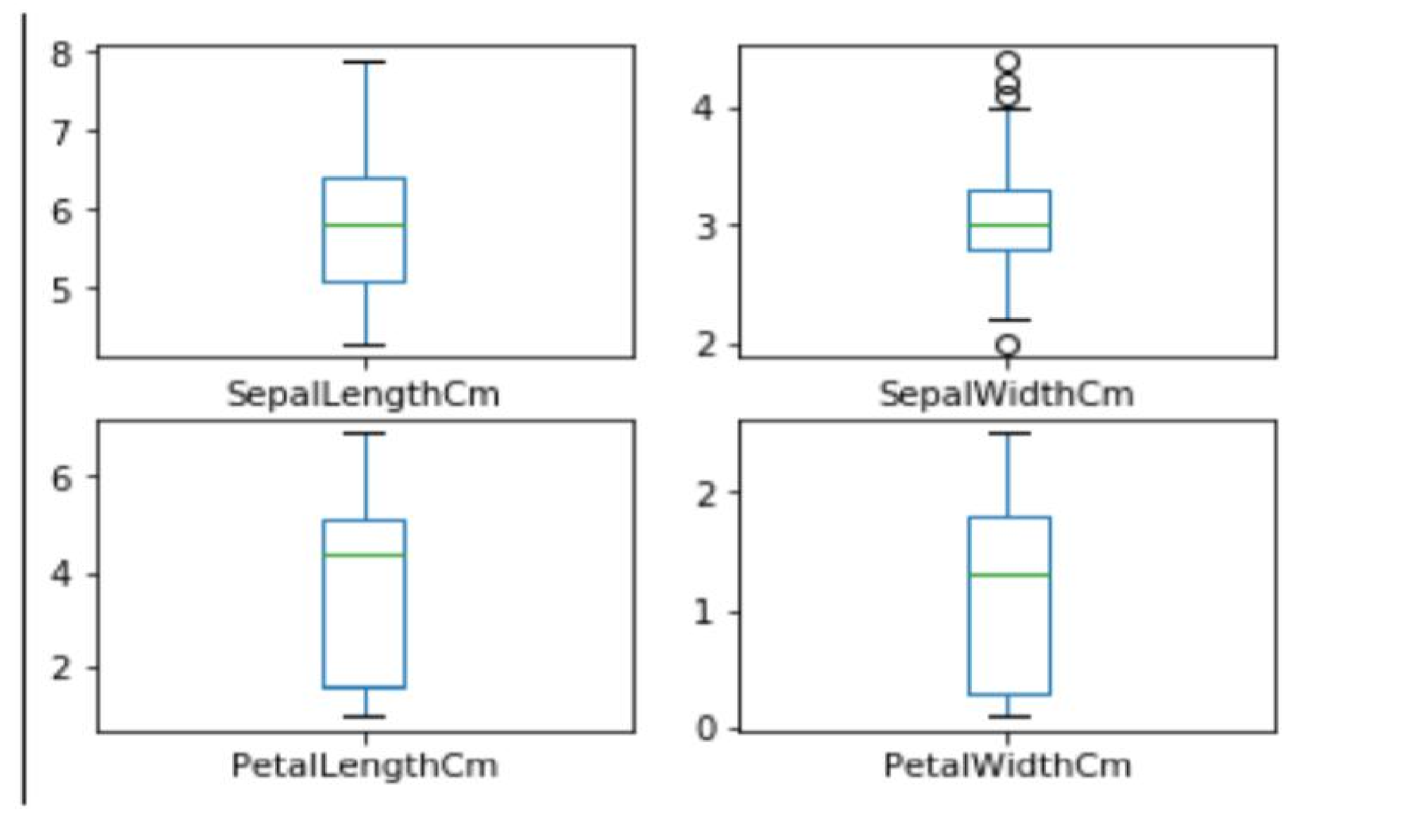
# Data Analysis :

1. Descriptive statistics- SD, Min, Max etc .

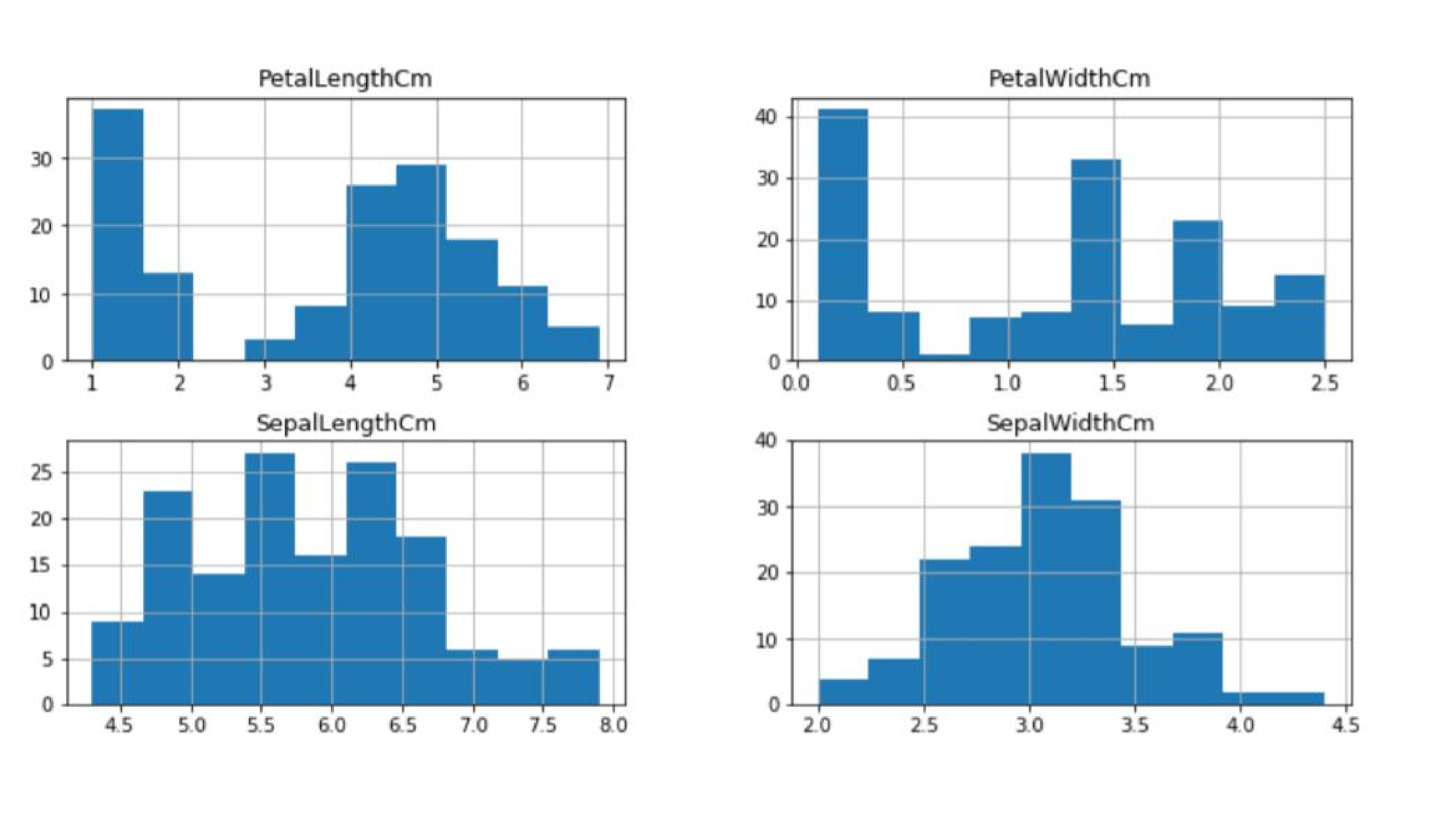
2. Class Distribution (Species counts are balanced or imbalanced) – Balanced.

3. Univariate Plots:- Understand each attribute better.

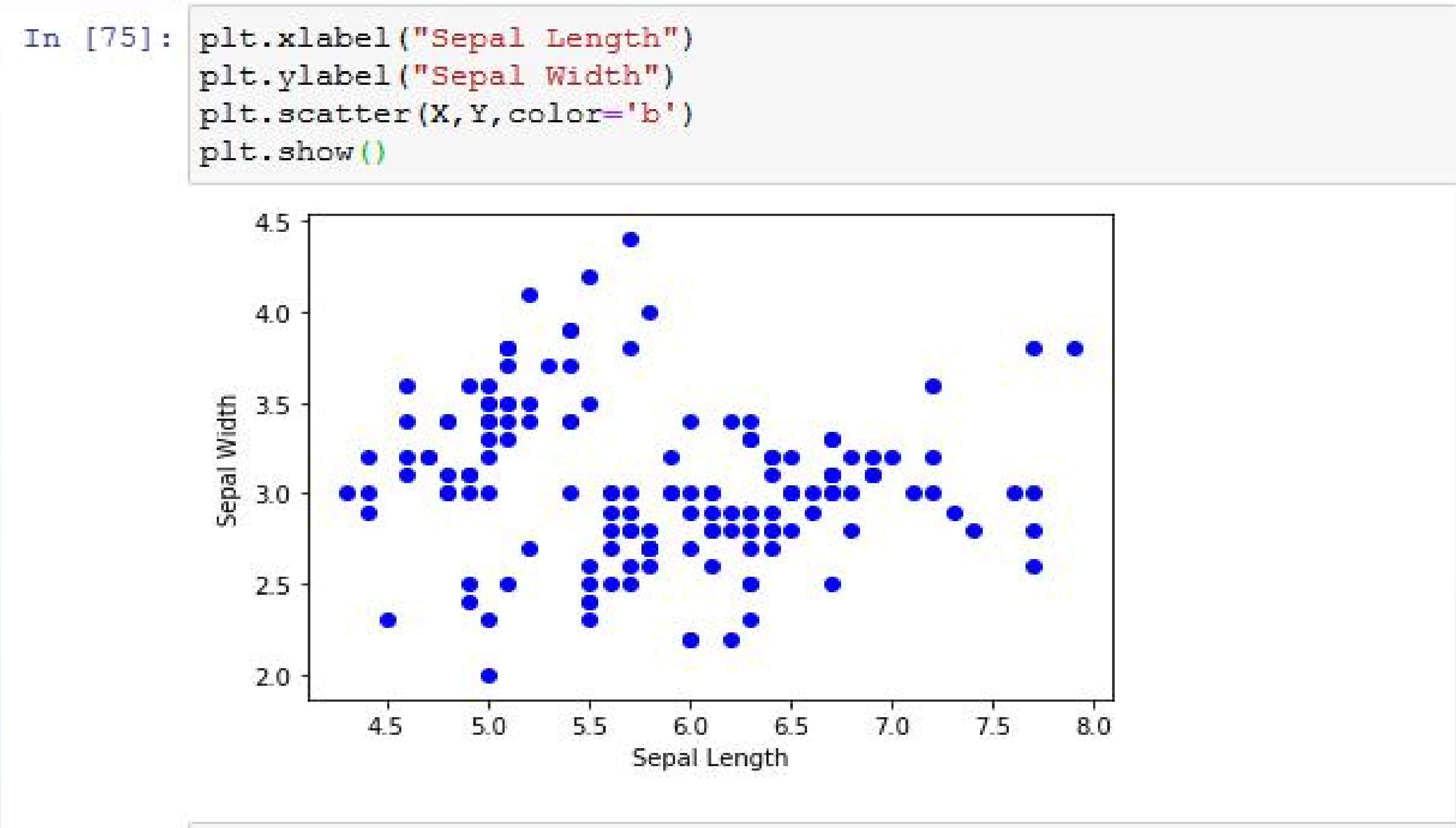
# Box and whisker plots(Give idea about distribution of input attributes)



# Plotting Histogram:



# Plotting Scatter Graph Between Sepal Length and Sepal Width:



```
In [75]: plt.xlabel("Sepal Length")
         plt.ylabel("Sepal Width")
         plt.scatter(X,Y,color='b')
         plt.show()
```

# Observation:

1. Using Sepal_Lenght & Sepal_Width features, we can only distinguish Setosa flower from others.

2. Seperating Versicolor & Virginica is much harder as they have considerable overlap.

3. Hence, Sepal_Lenght & Sepal_Width features only work well for Setosa.

# Implementation of Machine Learning.

# Steps to implement Machine Learning

1. Import Library

2. Analyze Data

3. Spliting the Data Set into train and test

4. Chossing right algorithm for training model

5. Test the algorithm with test data.

# Algorithms Used:

1. Logistic Regression

2. Support Vector Machine

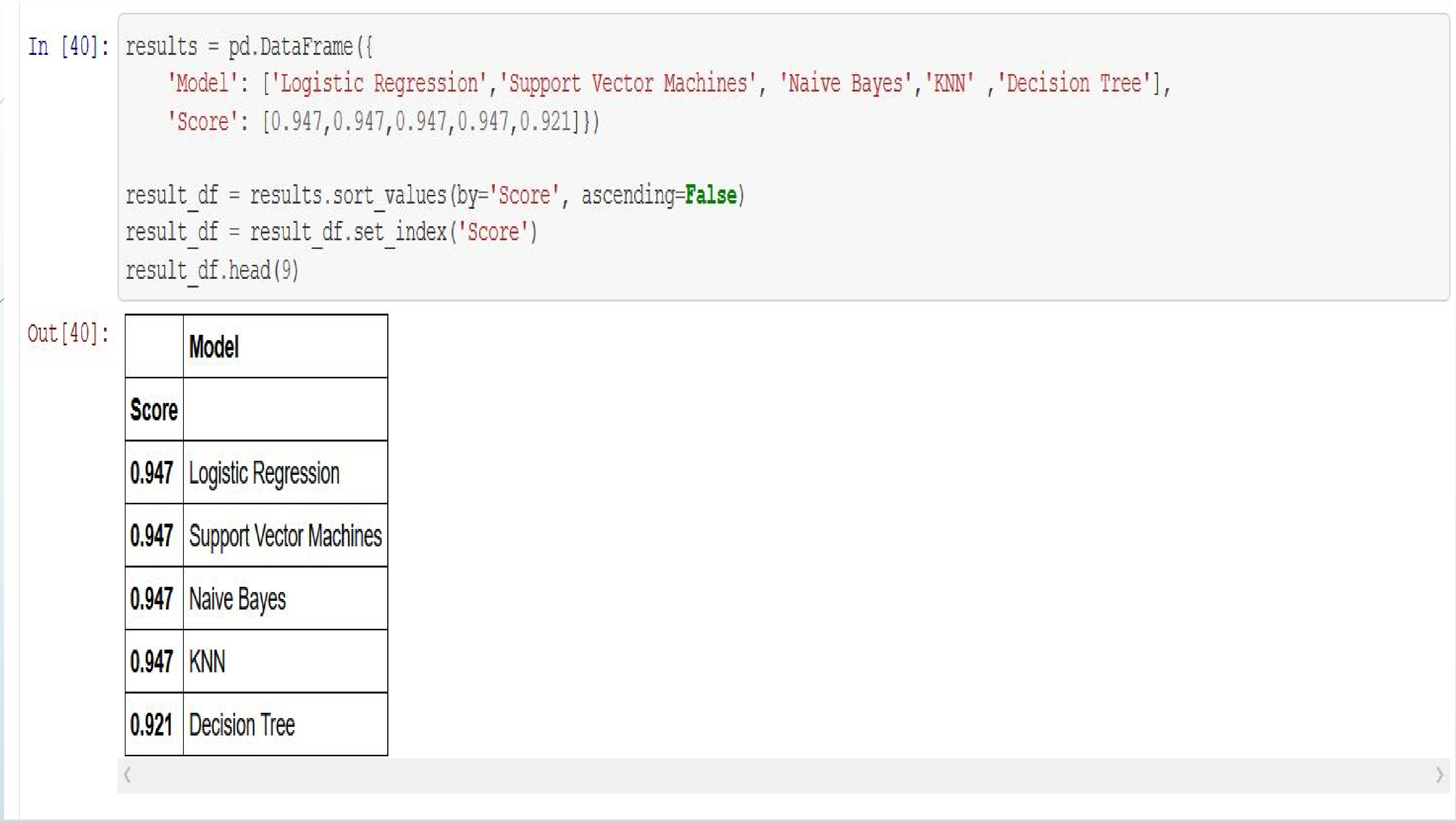
3. Classification and Regression Tree(CART)

4. Gaussion Naive Bayes(NB)

5. K-Nearest Neighbour(KNN)

6. Deision Tree

# Final Evaluation Of All Models:



In [40]:

```python
results = pd.DataFrame({
    'Model': ['Logistic Regression','Support Vector Machines', 'Naive Bayes','KNN' ,'Decision Tree'],
    'Score': [0.947,0.947,0.947,0.947,0.921]})

result_df = results.sort_values(by='Score', ascending=False)
result_df = result_df.set_index('Score')
result_df.head(9)
```

Out[40]:

| | **Model** |
|---|---|
| **Score** | |
| **0.947** | Logistic Regression |
| **0.947** | Support Vector Machines |
| **0.947** | Naive Bayes |
| **0.947** | KNN |
| **0.921** | Decision Tree |

THANK
YOU