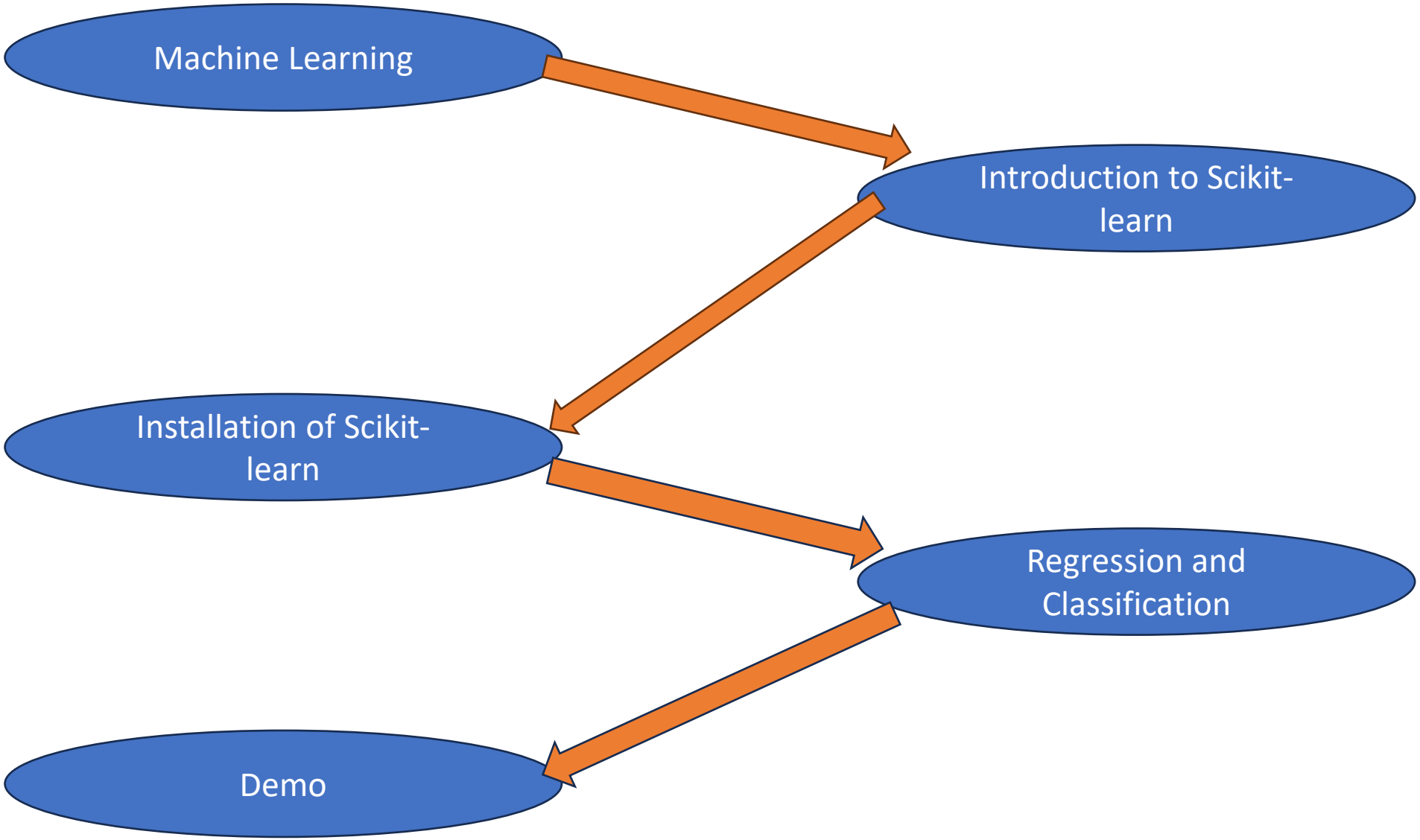
Lab: Scikit-Learn

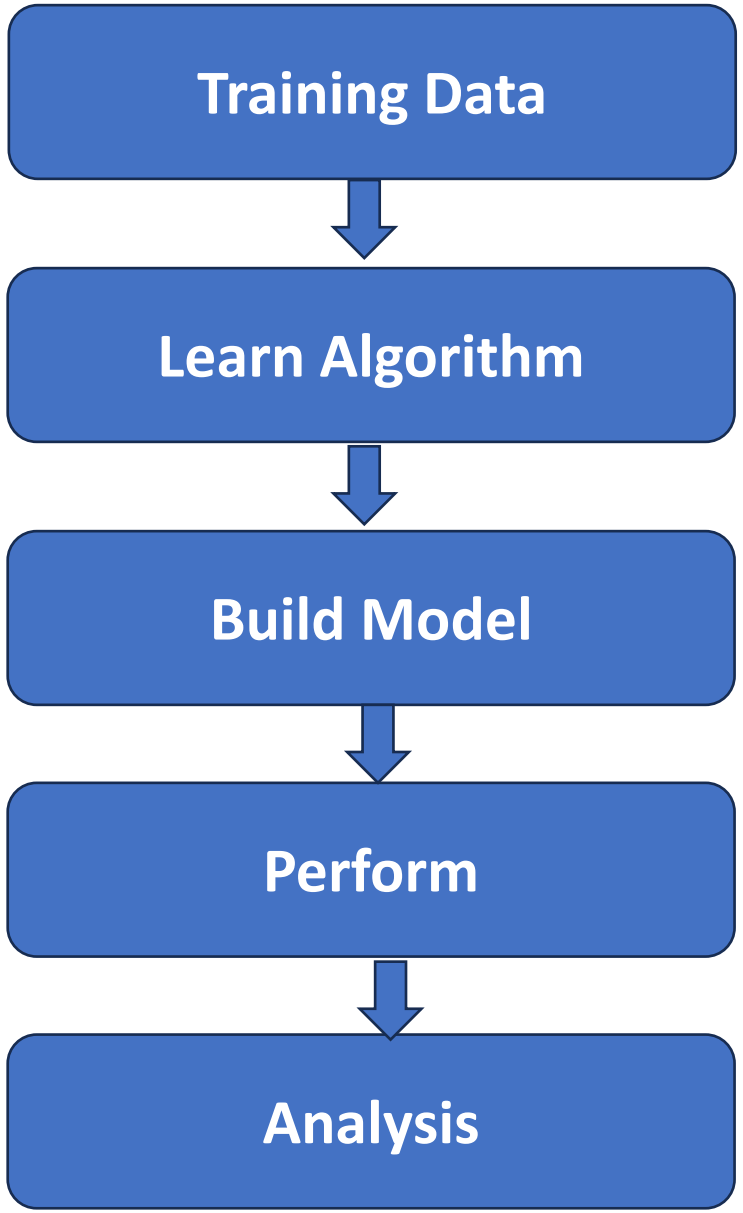
Topic Covered



Machine Learning

- It is a type of artificial intelligence that allow computers learn from data to make decisions or predictions. It involves training models on data to identify patterns and improve over time.
- 5 points to understand machine learning:
  - **Learning from Data**: Machines improve their performance by learning patterns from data.
  - **Types:** It includes supervised, unsupervised, and reinforcement learning methods.
  - **Training**: Models are trained on data to make accurate predictions or decisions.
  - **Prediction**: After training, models predict outcomes for new, unseen data.
  - **Adaptation**: Models improve over time as they are exposed to more data.

Diagram


Training Data
Learn Algorithm
Build Model
Perform
Analysis

Types of Machine Learning

01 Supervised

This is a process of an algorithm learning from the training dataset.

02 Unsupervised

This is a process where a model is trained using an information which is not labelled.

03 Reinforcement

Reinforcement learning is learning by interacting with a space or an environment.

Introduction to Scikit-learn

Library for Machine Learning: Scikit-learn is a popular Python library for machine learning, offering simple and efficient tools for data analysis.

Algorithms: It provides a wide range of algorithms for classification, regression, clustering, and dimensionality reduction.

Modelling Tools: Scikit-learn includes tools for model selection, evaluation, and data preprocessing.

Integration: It integrates well with other scientific libraries like NumPy, SciPy, and matplotlib.

Installation of Scikit-learn

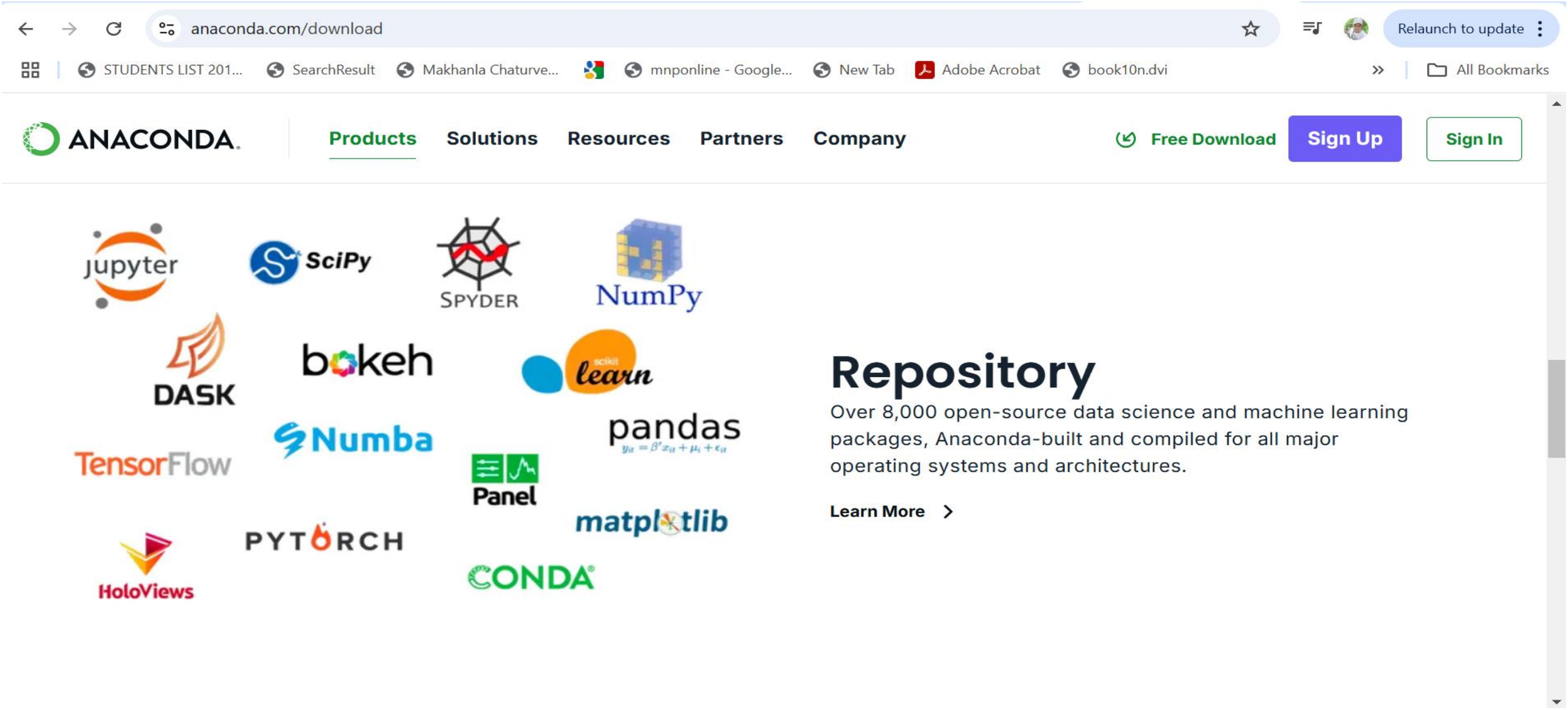
- Command:

```
Pip install scikit-learn
```

- Or

```
conda install scikit-learn
```

Installation from web



Import model

- Command:

```
from sklearn.family import Model
```

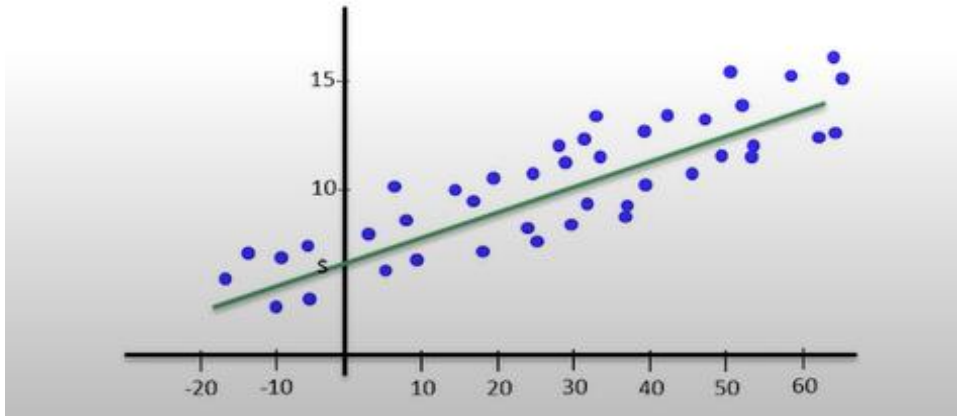
- Or

```
from sklearn.linear_model import LinearRegression
```

Regression and Classification

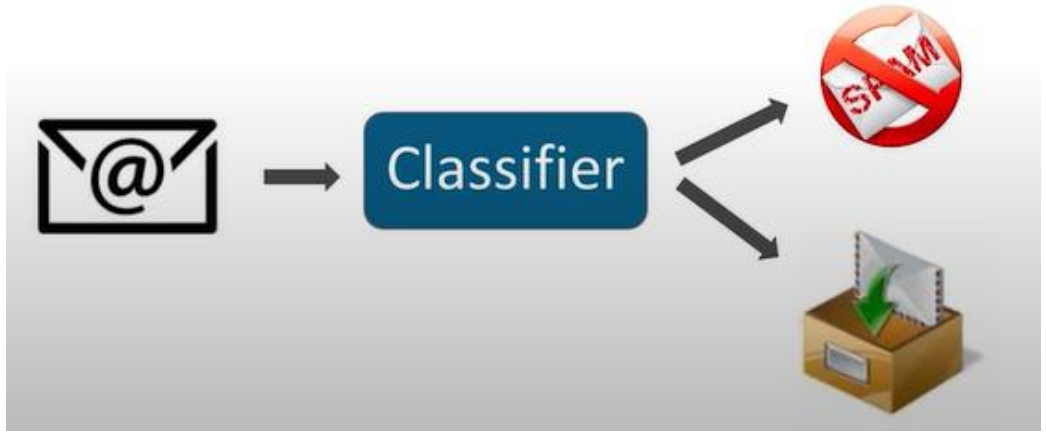
Regression

Regression is the prediction of a numeric value and often takes input as a continuous value.
Example: Salary Prediction, Revenue and House Price Prediction



Classification

Classification is problem identifying to which set of categories a new observation belong.
Example: spam Mail, Election prediction.



Dataset: IRIS

- The dataset consists of 50 samples from three species of Iris- Setosa, Virginica and versicolor.
- Four features were measured from each sample: Length and the width of the sepals and petals, in centimeters.
- Dataset link: https://www.kaggle.com/datasets/uciml/iris?select=Iris.csv