

# Popular libraries for ML & data science

List of all major libraries:

- Scikit-Learn
- TensorFlow
- PyTorch
- Keras
- XGBoost
- LightGBM
- Pandas
- NumPy
- Matplotlib
- Seaborn
- Hugging Face  
transformers

# NUMPY IN PYTHON

Numpy is a Python library for scientific computing that is widely used in the field of machine learning. It provides a high performance multidimensional array object, as well as tools for working with those arrays. In machine Learning

Numpy is often used to store and manipulate large amounts of data that are used as input to ML models, as well as to perform mathematical operations on this data in order to prepare it for use in machine learning algorithms.

Numpy can be imported in Python:

# Matplotlib, Pandas

Matplotlib is a comprehensive library for creating static, animated and interactive visualizations in Python.

```
import matplotlib.pyplot as plt
```

Python Command

## Pandas

Pandas is a fast powerful, flexible & easy ~~way~~ to use open source data

analysis & manipulation, built on top of Python Programming language.

It usually stores data in tabular format.

```
import Pandas as pd
```

Pandas helps with data manipulation, data cleaning and data analysis

making it easier to prepare data for model training & evaluation.

## Learn in Python

Scikit-Learn is a machine learning library for Python. It has a range of tools for building and training models.

Scikit-Learn has a focus on supervised learning and provides a range of algorithms for classification, regression and clustering.

It also has tools for feature selection, dimensionality reduction and model eval-

Scikit-learn is a popular choice for beginners for its simple syntax.

```
from sklearn.preprocessing import  
OneHotEncoder
```

```
from sklearn.model_selection import  
train_test_split,
```

```
from sklearn.linear_model import
```

# Tensorflow in Python

Tensorflow is a ml library made by google, it is used for building, training and deploying machine learning models, and has a strong focus on deep learning.

Tensorflow has a set of tools for building & training neural networks including support for CNN's and LSTM networks ; it is very efficient allowing developers to build models that could run on CPU's, GPU's and TPU's.

```
import tensorflow as tf
```

## Keras in Python

Keras is a ml library which is built on tensorflow. It is designed to make building & training deep learning models.

Keras provides a range of

# PyTorch in Python

PyTorch is a machine learning library for Python made by Facebook. It is made for a fast & flexible & it has a strong focus on deep learning.

PyTorch can also be easily modified while training making it well suited for large & complex models.

It is widely used in research & production level ml applications.

[import torch]

Research-focussed

PyTorch

Industry-focussed

vs. Tensorflow

Pros	Cons	Pros & Cons	Backward compatibility issues
Simplicity	limited	various	Tensorflow's backward compatibility issues
Flexibility	visualization options	Platform support	Tensorflow's performance & usability issues
Native integration	limited model serving	Very comprehensive tool	Tensorflow's learning curve
Easy			

## Tensorflow

- Good for scalability and real world-use in applications
- Deploying in real world situations

Tensorflowhub

1300 models

## PyTorch

- Easier for researchers as it allows for flexibility & creativity
- ChatGPT uses PyTorch as it is more for research

## PyTorch

User friendly  
Dynamic computation graphs

## Tensorflow

- More libraries
- more time to learn

Seaborn is a data visualization library based on matplotlib lib.

It has visually appealing graphs for formal presentations.

Import Seaborn as sns

xgboost

xgboost stands for extreme gradient boosting, it uses

a gradient boosting algorithm to solve data

science problems.

LightGBM

LightGBM is a ensemble learning framework using a

# Vocabulary for ML & DATA SCIENCE

ablation - A technique for evaluating the importance of a feature or component by temporarily removing it from a model.

Ex. Remove 10<sup>th</sup> weight and accuracy drops from 88% to 55% then the weight is important

classification is a part of supervised learning which inputs can be separated easily into two classes.

A decision tree is often used for classification problems.

# Sectors in ML and Jobs

## Natural Language Processing

This is the field of AI that deals with human language. The challenge is that human speech is confusing.

First, it converts unstructured language data into a form that can be understood by a computer, then it is passed through an algorithm that interprets the meaning of the words.