



# **UNIVERSIDAD TECNOLÓGICA DE SAN LUIS RIO COLORADO**

**Librerías investigación**

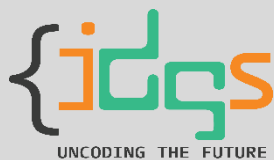
**MTRO. AURELIO FLORES**

**ALUMNO: VICTOR MANUEL GALVAN COVARRUBIAS**

**ING. EN DESARROLLO Y GESTIÓN DE SOFTWARE**

San Luis Río Colorado, Sonora

Septiembre, 2022



## A. Pandas



Pandas is a Python package that provides fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python.

conda install pandas

Things that pandas do:

- Easy handling of missing data (represented as NaN, NA, or NaT) in floating point as well as non-floating point data.
- Size mutability: columns can be inserted and deleted from DataFrame and higher dimensional objects.
- Automatic and explicit data alignment: objects can be explicitly aligned to a set of labels, or the user can simply ignore the labels and let Series, DataFrame, etc. automatically align the data for you in computations
- Powerful, flexible group by functionality to perform split-apply-combine operations on data sets, for both aggregating and transforming data
- Make it easy to convert ragged, differently-indexed data in other Python and NumPy data structures into DataFrame objects
- Intelligent label-based slicing, fancy indexing, and subsetting of large data sets
- Intuitive merging and joining data sets
- Flexible reshaping and pivoting of data sets
- Hierarchical labeling of axes (possible to have multiple labels per tick)
- Robust IO tools for loading data from flat files (CSV and delimited), Excel files, databases, and saving/loading data from the ultrafast HDF5 format
- Time series-specific functionality: date range generation and frequency conversion, moving window statistics, date shifting and lagging

## B. Numpy



NumPy is a Python library used for working with arrays.

It also has functions for working in domain of linear algebra, fourier transform, and matrices.

NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

NumPy stands for Numerical Python.

Numpy makes many mathematical operations used widely in scientific computing fast and easy to use, such as:

- Vector-Vector multiplication
- Matrix-Matrix and Matrix-Vector multiplication
- Element-wise operations on vectors and matrices (i.e., adding, subtracting, multiplying, and dividing by a number)
- Element-wise or array-wise comparisons
- Applying functions element-wise to a vector/matrix (like pow, log, and exp)
- A whole lot of Linear Algebra operations can be found in NumPy.linalg
- Reduction, statistics, and much more

## C. Scipy



SciPy in Python is an open-source library used for solving mathematical, scientific, engineering, and technical problems. It allows users to manipulate the data and visualize the data using a wide range of high-level Python commands. SciPy is built on the Python NumPy extension. SciPy is also pronounced as “Sigh Pi.”

Why use SciPy

- SciPy contains varieties of sub packages which help to solve the most common issue related to Scientific Computation.
- SciPy package in Python is the most used Scientific library only second to GNU Scientific Library for C/C++ or Matlab's.
- Easy to use and understand as well as fast computational power.
- It can operate on an array of NumPy library.
- SciPy is built in top of the NumPy
- SciPy module in Python is a fully-featured version of Linear Algebra while Numpy contains only a few features.
- Most new Data Science features are available in Scipy rather than Numpy.

## D. Scikit-learn



Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistency interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

### Features

- Datasets. Scikit-learn comes with several inbuilt datasets such as the iris dataset, house prices dataset, diabetes dataset, etc.
- Data Splitting.
- Linear Regression.
- Logistic Regression.
- Decision Trees.
- Random Forest.
- XG Boost.
- Support Vector Machines(SVM)

## E. Matplotlib



Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

- Create publication quality plots.
- Make interactive figures that can zoom, pan, update.
- Customize visual style and layout.
- Export to many file formats.
- Embed in JupyterLab and Graphical User Interfaces.
- Use a rich array of third-party packages built on Matplotlib.

## F. TensorFlow



TensorFlow gives us an interactive multiplatform programming interface which is scalable and much stable when compared to other **deep learning** libraries available, which are still very experimental.

### Features of Tensorflow

- a. Responsive Construct.
- b. Flexible.
- c. Easily Trainable.
- d. Parallel Neural Network Training.
- e. Large Community.
- f. Open Source.
- g. Feature Columns.
- h. Availability of Statistical Distributions.