  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
TELEINFORMÁTICA I - **​GRUPO 020-82**

**ESTUDIANTE:**

DAVID FELIPE VEGA - 20182020033

**REPOSITORIO GIT:**

https://github.com/dfvegas11/Teleinformatica1

**PROFESOR:**

ALBERTO ACOSTA LOPEZ

**BOGOTA D.C**

**SCIKIT – LEARN**

La ciencia de datos cada vez es más fuerte dentro del mundo de la programación. La cantidad de procesos y el manejo de grandes cantidades de datos pueden definirse en diferentes lenguajes de programación, sin embargo, Python es uno de los más populares.



Ilustración 1. Logo de Scikit-learn.

Scikit-learn es una herramienta para el desarrollo Data Science en Python. Siendo una de las librerías gratuitas de Python cuenta con algoritmos de clasificación, regresión, clustering y reducción de dimensionalidad. Además, presenta la compatibilidad con otras librerías numéricas y científicas de Python como Numpy, SciPy y Matplotlib.

Gracias a la gran cantidad de algoritmos y utilidades de Scikit-learn la han convertido en la herramienta básica para empezar a programar y estructurar los sistemas de análisis datos y modelado estadístico. Estos algoritmos utilizan aplicaciones externas como Pandas o PyBrain para mejorar la depuración de las estructuras de datos y obtener resultados mucho más aproximados.

**HISTORIA**

Scikit-learn se inició en 2007 como un proyecto Google Summer of Code de David Cournapeau, su nombre viene de la idea que se trata de un una extensión auxiliar desarrollada y distribuida independientemente de SciPy. Ese mismo año Matthieu Brucher trabajó en este proyecto como parte de su tesis.

En 2010, Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort y Vincent Michel (todos pertenecientes a INRIA) asumieron el liderazgo del proyecto e hicieron el primer lanzamiento público el 1 de febrero de 2010. Desde entonces, han aparecido varios lanzamientos después de un ciclo de 3 meses y una comunidad internacional ha estado liderando el desarrollo.

Las siguientes personas son actualmente los colaboradores principales para el desarrollo y mantenimiento de scikit-learn:

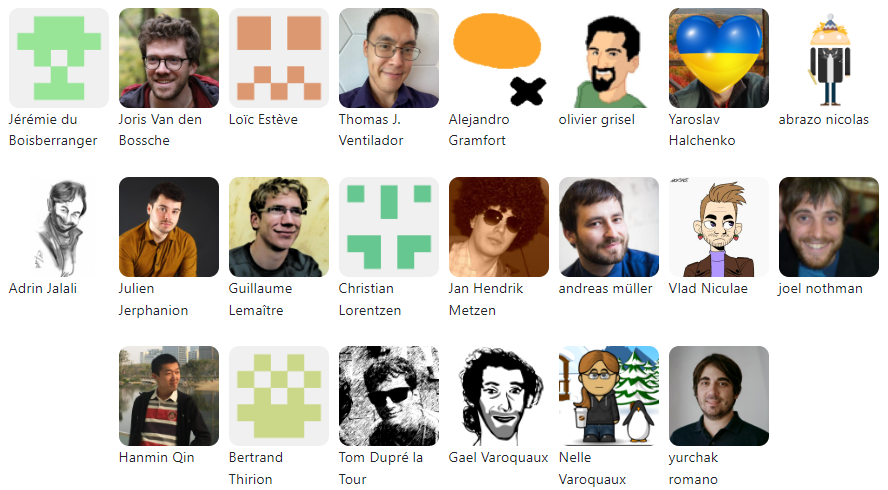


Ilustración 2. Actuales colaboradores de Scikit-learn.

**INSTALACIÓN**

* **Windows y Mac OS:**

Para la instalación de Scikit-learn en Windows y Mac OS se debe tener instalada la versión de 64 bits de Python 3 que se puede encontrar en <https://www.python.org>. En Mac OS también se puede usar homebrew (*brew install python*)

Una vez Instalado Python, atreves de su gestor de paquetes *pip* podremos instalar la biblioteca usando el siguiente comando.

pip install -U scikit-learn

Para comprobar su instalación se puede utilizar el comando

python -m pip show scikit-learn #para ver la versión instalada

* **Linux:**

Para la instalación de Scikit-learn en Linux debemos instalar python3 y python3-pip utilizando el administrador de paquetes de la distribución de Linux.

Una vez Instalado Python, atreves de su gestor de paquetes *pip* podremos instalar la biblioteca usando el siguiente comando.

pip3 install -U scikit-learn

Para comprobar su instalación se puede utilizar el comando

python3 -m pip show scikit-learn #para ver la versión instalada

* **Conda:**

Otra opción para la instalación de Scikit-learn es por medio del gestor de paquetes Conda, para esto se debe instalar por medio de los instaladores de Anaconda o miniconda (<https://docs.conda.io/projects/conda/en/latest/user-guide/install/>) y luego ejecutamos el siguiente comando.

conda install -c anaconda scikit-learn

Para comprobar su instalación se puede utilizar el comando

conda list scikit-learn #para ver la versión instalada

La página oficial de Scikit-learn recomienda utilizar un entorno virtual con la finalidad de evitar posibles conflictos con otros paquetes instalados.

A continuación, se muestran la versión mínima de las dependencias de Scikit-learn junto con su propósito.

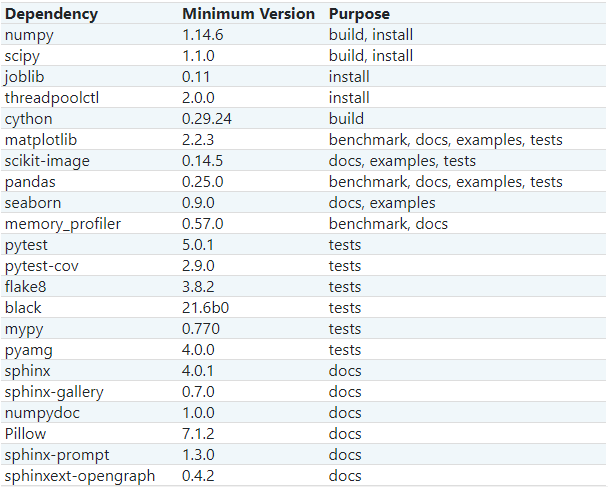


Ilustración 3. Dependencias de Scikit-learn.

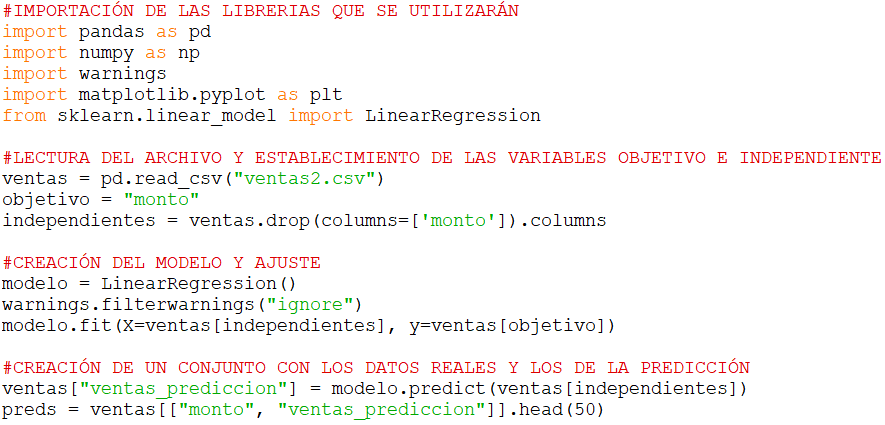
**EJEMPLO**

Para el siguiente ejemplo usaremos un archivo CSV (un archivo separado por comas) con información sobre ventas de cualquier tipo de tienda.

Dentro de este archivo encontramos los campos:

* Edad: se refiere a la edad del cliente.
* Cantidad: se refiere a la cantidad de personas que iban con esa persona.
* Vehículo: indica si llegó o no en un vehículo particular.
* Pago: se refiere a su forma de pago (efectivo, tarjeta de crédito, débito o vales).
* Monto: que indica cuánto compró.

Crearemos un modelo sencillo al que le indiquemos edad, cantidad de acompañantes, vehículo y forma de pago y él nos prediga cuánto va a comprar esa persona.



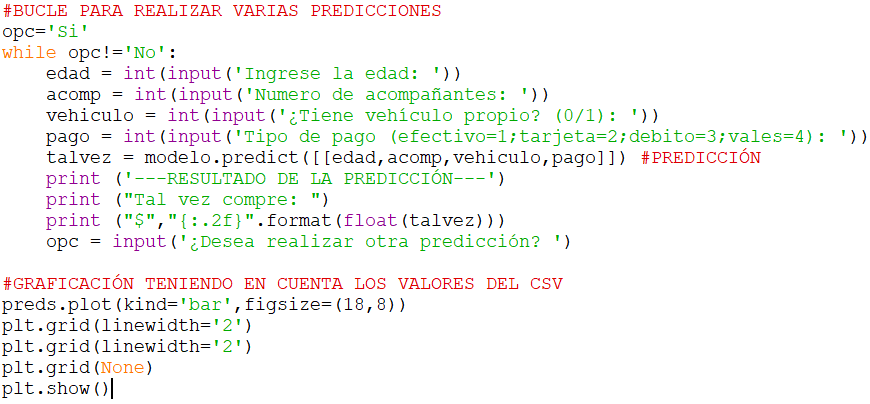


Ilustración 4. Código ejemplo de utilización Scikit-learn.

**REFERENCIAS**

* Scikit-Learn, herramienta básica para el Data Science en Python. (2018, agosto 6). Máster en Data Science. <https://www.master-data-scientist.com/scikit-learn-data-science/>
* Scikit-learn. (2020). En Wikipedia, la enciclopedia libre. <https://es.wikipedia.org/w/index.php?title=Scikit-learn&oldid=130746122>
* Scikit-learn: Machine learning in Python—Scikit-learn 1.0.2 documentation. (s. f.). Recuperado 3 de mayo de 2022, de <https://scikit-learn.org/stable/>