

PySciT

A Python Scikit-learn Pipelines Toolkit

Manual de usuario

Leticia Quiñones Rodríguez

Septiembre de 2023

Índice

PySciT	4
Instalación	4
Pantallas principales	5
Home	5
Log	5
Configuration List	6
Help	6
Manage Datas Menu	7
Manage Pipelines Menu	7
Crear una tabla con datos locales	8
Create a new Data	8
Página de una tabla	10
Crear un pipeline	10
Create a new Pipeline	11
Página de un pipeline	11
Modificar un pipeline	12
Modificar steps de un pipeline	13
Ejecutar un pipeline	14

PySciT

PySciT (Python Scikit-learn Pipelines Toolkit) es una aplicación web con el objetivo de permitir el diseño arquitectónico de pipelines personalizables para su uso en el aprendizaje automático.

A través de una interfaz sencilla, ofrece al usuario herramientas que le permitan gestionar, construir y ejecutar pipelines personalizados de aprendizaje automático de forma fácil y accesible. Está desarrollada en Python y utiliza librerías como Scikit-Learn o Pandas entre otras.

Instalación

El código de este proyecto se puede encontrar en el siguiente enlace a GitHub: <https://github.com/lquirod/PySciT>. Para poder descargarlo y utilizarlo se siguen los siguientes pasos en un terminal:

1. Clonar el proyecto con git:
\$ git clone https://github.com/lquirod/PySciT
2. Crear entorno virtual dentro del proyecto y activarlo:
\$ cd PySciT/
\$ python3 -m venv venv/
\$ source venv/bin/activate

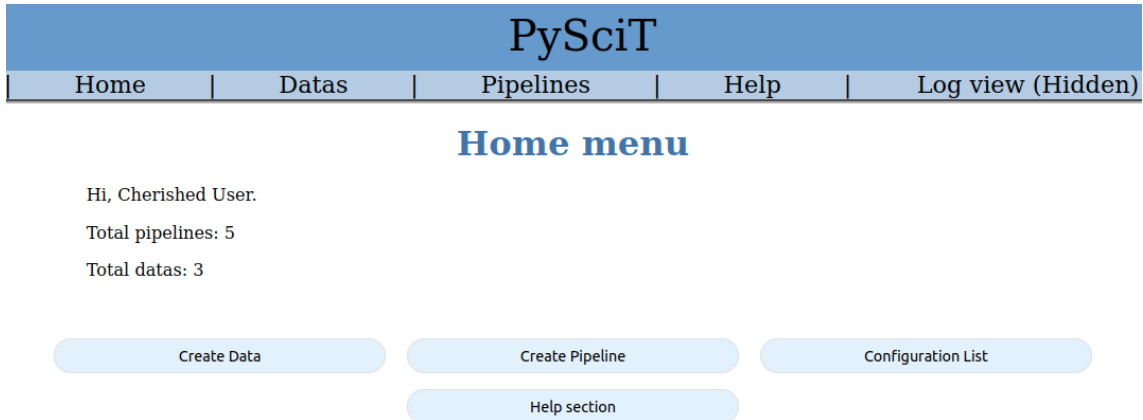
Cuando el entorno esté activado se indicará en nuestro terminal como

```
(venv) myuser:path$
```

3. Instalar los paquetes necesarios:
\$ pip3 install -r requirements.txt
4. Instalar la aplicación web de Flask:
\$ pip install -e .
5. Ejecutar la aplicación localmente:
\$ flask --app PySciT_web run
6. Abrir en el navegador el enlace indicado. Suele ser:
\$ http://127.0.0.1:5000/

Pantallas principales

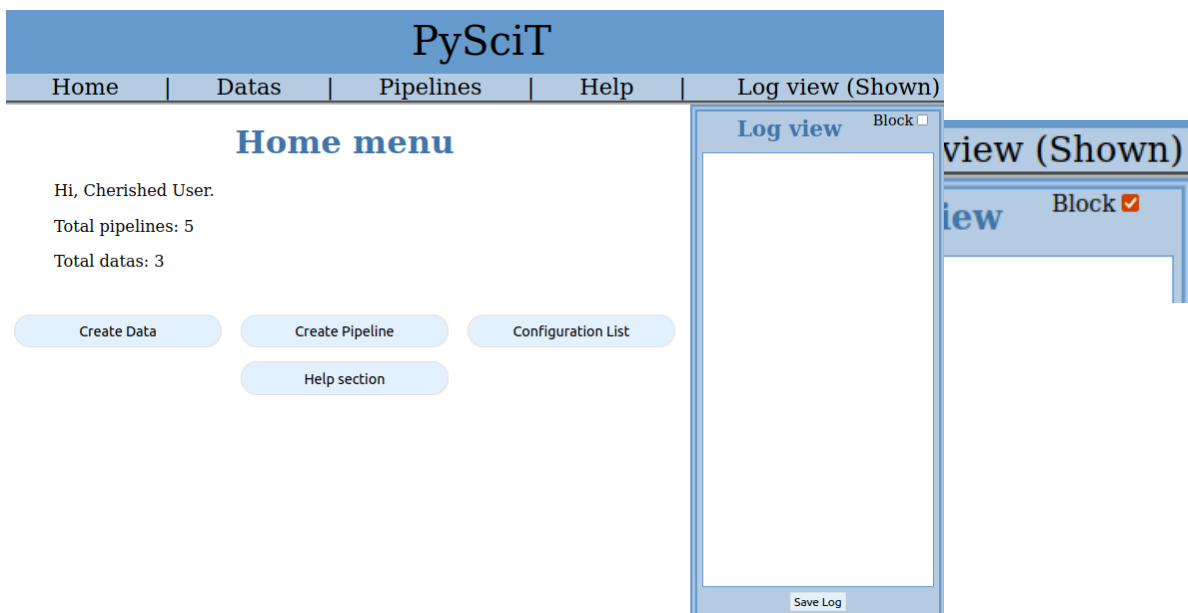
Home



Home es la página inicial. Muestra un resumen general (total de pipelines y datas) y enlaces a páginas de interés. Se puede observar un menú navegable superior con las páginas principales y el Log.

Log

El Log es una sección desplegable en el lateral derecho donde se muestra el historial de las acciones realizadas por el usuario. Al pulsar "Log view" se mostrará y se ocultará al pulsarlo de nuevo o cambiar de página. Si deseamos que no se oculte al cambiar de página, podemos marcar la casilla al lado de "Bloc".



Se puede descargar el contenido del Log pulsando “Save Log”.

Configuration List

En esta página se muestran las transformaciones y los algoritmos disponibles que tiene el sistema en ese momento.

PySciT

Home | Datas | Pipelines | Help | Log view (Hidden)

← Back Home

Configuration List

This is the list of transformations available add them as steps of a Pipeline

- MinMax_Scaling

And this is the list of the algorithms available to apply them on the Data items or add them as steps of a Pipeline

- Linear_Regression
- KNeighbors_Classifier
- Support_Vector_Classification
- Random_Forest_Classifier

Help

PySciT

Home | Datas | Pipelines | Help | Log view (Hidden)

← Back Home

Help section

What's PySciT

PySciT, a Python Scikit-learn Pipelines Toolkit, is a system a tool for the management, construction and execution of pipelines for Machine Learning.

The objective of this project is to develop a system that allows the architectural design of customizable pipelines for its use in machine learning. This tool has been designed and implemented for the management, construction and execution of pipelines in Python, using libraries such as Scikit-Learn or Pandas among others, through a simple interface to this end.

You can find the repository code here in gitHub: <https://github.com/lquirod/PySciT>

User manual

0 of 0

Automatic Zoom

En esta página se puede consultar este mismo manual de usuario.




Manage Datas Menu

PySciT				
Home	Datas	Pipelines	Help	Log view (Hidden)


[← Back Home](#)

Manage Datas Menu

Total Datas: 3Create a new data +

Nº	Name	Columns	Rows	Access
0	New Data	None	0	
1	Example database	['Name', 'Birth year', 'Birth month', 'Likes']	10	
2	Test data 1	['X_train1', 'X_train2', 'y_train', 'Xtrain1y2']	4	

Muestra la lista de las tablas (Data) de datos guardados indicando su nº, su nombre, los campos que posee y el total de registros de cada uno. Se puede crear una nueva tabla en “Create a new data +” en el lado superior derecho.

El icono  lleva a la página de previsualización de la tabla correspondiente.






Manage Pipelines Menu

PySciT				
Home	Datas	Pipelines	Help	Log view (Hidden)

[← Back Home](#)

Manage Pipelines Menu

Total Pipelines: 5Create a new pipeline +

Nº	Name	Algorithm	Steps	Access
0	MyNewPipe	Linear_Regression	6	
1	Linear_Regression	Linear_Regression	1	
2	Support_Vector_Classification	Support_Vector_Classification	1	
3	KNeighbors_Classifier	KNeighbors_Classifier	1	
4	Random_Forest_Classifier	Random_Forest_Classifier	1	

Muestra la lista de los pipelines creados indicando su nº, su nombre, el algoritmo que posee y el total de steps (o pasos) de cada uno. Se puede crear un nuevo pipeline en “Create a new pipeline +” en el lado superior derecho.

El icono  lleva a la página de detalles del pipeline correspondiente.

Crear una tabla con datos locales

Se puede crear una nueva tabla (Data) desde los botones “Create Data” del menú principal o “Create a new data +” en el lado superior derecho del *Manage Datas Menu*. Ambos nos llevará a la página de creación de tabla.

PipelinesHelpLog view (Hidden)

Manage Datas Menu

PySciT

HomeDatsPipelines

Home menu

Hi, Cherished User.
Total pipelines: 5
Total datas: 3

Create DataCreate PipelineHelp section

Create a new data +

Columns	Rows	Access
None	0	
Birth year', 'Birth month', 'Likes']	10	
1', 'X_train2', 'y_train', 'Xtrain1y2']	4	

Create a new Data

El usuario puede cargar una tabla con datos a partir de un archivo .csv local. Para ello debe pulsar en “Browse...” y seleccionar el archivo .csv.

PySciT

HomeDatsPipelinesHelpLog view (Hidden)

Create a new Data

Back to Data Menu

File: No file selected. ☒ Data have name columns

Load a file first: .csv

No olvide indicar si el archivo contiene el nombre de los campos nombrados en la primera línea marcando el recaudo delante de “Data have name columns” (marcada por defecto) o no.

Create a new

File: ConNameCols.csv ☒ Data hav

it: .CSV

Una vez hecho, se pulsa el botón “Load”, que mostrará una previsualización de un máximo de 20 datos.

PySciT

Home
Datas
Pipelines
Help
Log view (Hidden)

Back to Data Menu

Create a new Data

File: No file selected.

☒ Data have name columns

☒ Id

☒ Name

☒ Site

0010900001sBs4xAAC	Burlington Textiles Corp of America	US-East	0010900001sBs4xAAC
0010900001sBs4zAAC	Dickenson plc	US-Midwest	0010900001sBs4zAAC
0010900001sBs4wAAC	Edge Communications	US-South	0010900001sBs4wAAC
0010900001sBs52AAC	Express Logistics and Transport	US-West	0010900001sBs52AAC
0010900001sBs56AAC	GenePoint	US-West	0010900001sBs56AAC
0010900001sBs50AAC	Grand Hotels & Resorts Ltd	US-Midwest	0010900001sBs50AAC

Preview file ConNameCols.csv:

- 20 rows (of 41)
- 12 columns

New data's name:

Select and rename the columns you want to save.

Ahora se puede cambiar el nombre de la tabla o seleccionar y renombrar los campos que se quieran guardar. Al terminar pulsar “Create Data”.

PySciT

Home
Datas
Pipelines
Help
Log view (Hidden)

Back to Data Menu

Create a new Data

File: No file selected.

☒ Data have name columns

☐ NameChange

☒ Site

☐ Id.1

AC	Burlington Textiles Corp of America	US-East	0010900001sBs4xAAC
AC	Dickenson plc	US-Midwest	0010900001sBs4zAAC
AC	Edge Communications	US-South	0010900001sBs4wAAC
AC	Express Logistics and Transport	US-West	0010900001sBs52AAC
AC	GenePoint	US-West	0010900001sBs56AAC
AC	Grand Hotels & Resorts Ltd	US-Midwest	0010900001sBs50AAC

Preview file ConNameCols.csv:

- 20 rows (of 41)
- 12 columns

New data's name:

Select and rename the columns you want to save.

Página de una tabla

Se muestra la tabla creada, la cual podremos modificar su nombre editando el recuadro y pulsando “Change name” o descargarlo de nuevo en “Download CSV”.

PySciT

Home | Datas | Pipelines | Help | Log view (Hidden)

Back to Data Menu

Data 3: Change name

Data preview (total: 41 rows)

	Id	NameChange	Site
0	0010900001sBs4xAAC	Burlington Textiles Corp of America	US-East
1	0010900001sBs4zAAC	Dickenson plc	US-Midwest
2	0010900001sBs4wAAC	Edge Communications	US-South
3	0010900001sBs52AAC	Express Logistics and Transport	US-West
4	0010900001sBs56AAC	GenePoint	US-West
5	0010900001sBs50AAC	Grand Hotels & Resorts Ltd	US-Midwest
6	0010900001sBs4yAAC	Pyramid Construction Inc.	EMEA
7	0010900001sBs51AAC	United Oil & Gas Corp.	US-East
8	0010900001sBs55AAC	United Oil & Gas Singapore	APAC

Data's name:

Change name

Download CSV

Crear un pipeline

Se puede crear un nuevo pipeline desde los botones “Create Pipeline” del menú principal o “Create a new pipeline +” en el lado superior derecho del *Manage Pipelines Menu*. Ambos nos llevará a la página de creación de tabla.

PySciT

Pipelines | Help | Log view (Hidden)

Manage Pipelines Menu

Create a new pipeline +

Name	Steps	Access
Regression	6	
Regression	1	
Classification	1	

Home menu

and User.

ies: 5

3

Create Data

Create Pipeline

Config

Help section

Create a new Pipeline

Para crear un pipeline hay que seleccionar uno de los algoritmos listados, poner un nombre a nuestro pipeline (opcional) y pulsar “Create pipeline”.

PySciT

Home | Datas | Pipelines | Help | Log view (Hidden)

Back to Pipeline Menu

Create a new Pipeline

Total Pipelines Types: 4

Algorithms	Select
Linear_Regression	<input type="radio"/>
KNeighbors_Classifier	<input type="radio"/>
Random_Forest_Classifier	<input type="radio"/>
Support_Vector_Classification	<input type="radio"/>

New pipeline's name:

Create pipeline

Página de un pipeline

Se muestra la información de un pipeline creado, el cual podremos modificar su nombre editando el recuadro y pulsando “Change name”.

PySciT

Home | Datas | Pipelines | Help | Log view (Hidden)

Back to Pipeline Menu

Pipeline 0: MyNewPipe

Total Steps: 2

Modify disabled

Pos	Steps' Name	Move	Delete
0	MinMax_Scaling	⬇	<input type="button" value="Delete"/>
1	Linear_Regression	⬆	<input type="button" value="Delete"/>

Pipeline's name:

MyNewPipe

Change name

Save Setup

Add Step

Parameters

Delete pipeline

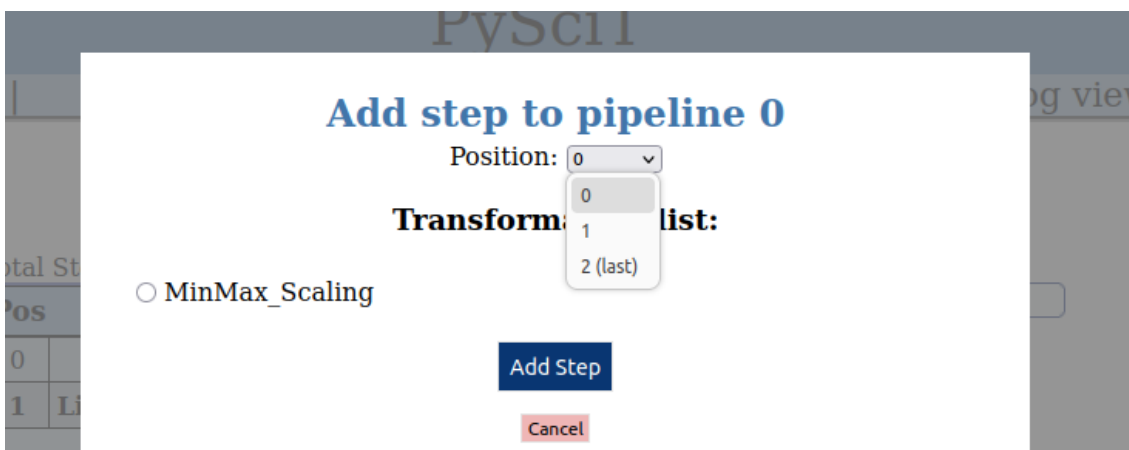
Pipeline operations

Fit	<input type="button" value="Operate"/>
Predict	<input type="button" value="Operate"/>
Score	<input type="button" value="Operate"/>

Modificar un pipeline

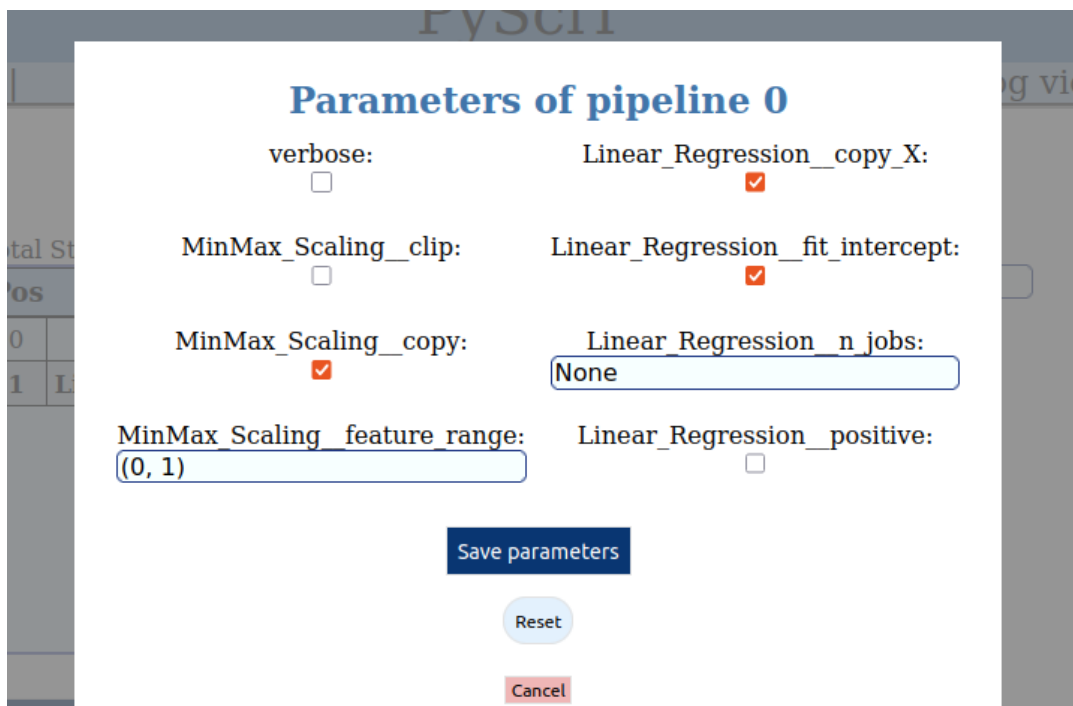
En la página de un pipeline podemos observar los siguientes botones:

1. Save Setup: Descarga un archivo con las configuraciones actuales del pipeline
2. Add Step: Añade un step al pipeline. Aparece un recuadro donde podremos seleccionar la transformación a añadir y su posición.



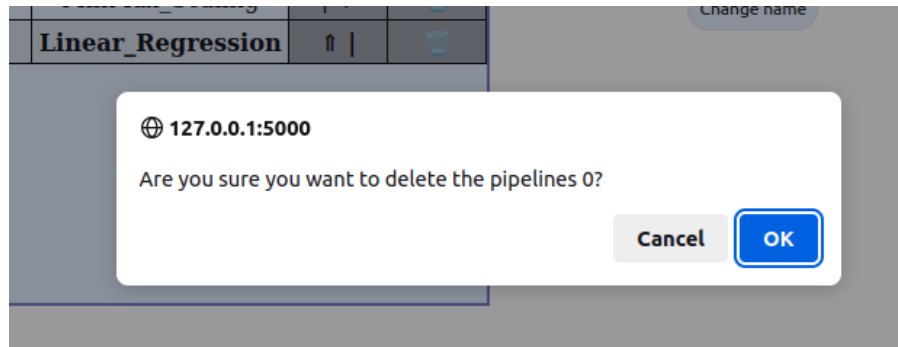
Pulsar "Add Step" para confirmar o "Cancel" para cancelar la operación.

3. Parameters: Aparece un recuadro donde podremos visualizar y modificar los parámetros del pipeline seleccionado.



Pulsar “Save parameters” para confirmar, “Reset” para resetear o “Cancel” para cancelar la operación.

4. Delete pipeline: Elimina un pipeline. La operación solo se hará efectiva después de que el usuario confirme la notificación.



Modificar steps de un pipeline

Para modificar los steps de un pipeline primero debemos pulsar el botón “Modify disabled” en la parte superior derecha de la tabla de steps, el cual cambiará a “Modify enabled” y nos permitirá realizar las siguientes operaciones:

Back to pipeline Menu

Pipeline 0: MyNewPipe

Total Steps: 3

Modify enabled

Pos	Steps' Name	Move	Delete
0	MinMax_Scaling	↓	
1	MinMax_Scaling	↑ ↓	
2	Linear_Regression	↑	

1. Mover un step: Para mover de posición un step sólo hay que pulsar la flecha en la dirección deseada que está en la misma fila del step. (Ej mover step 2 una posición anterior pulsando ↑)

Total Steps: 3

Modify enabled

Pos	Steps' Name	Move	Delete
0	MinMax_Scaling	↓	
1	MinMax_Scaling	↑ ↓	
2	Linear_Regression	↑	

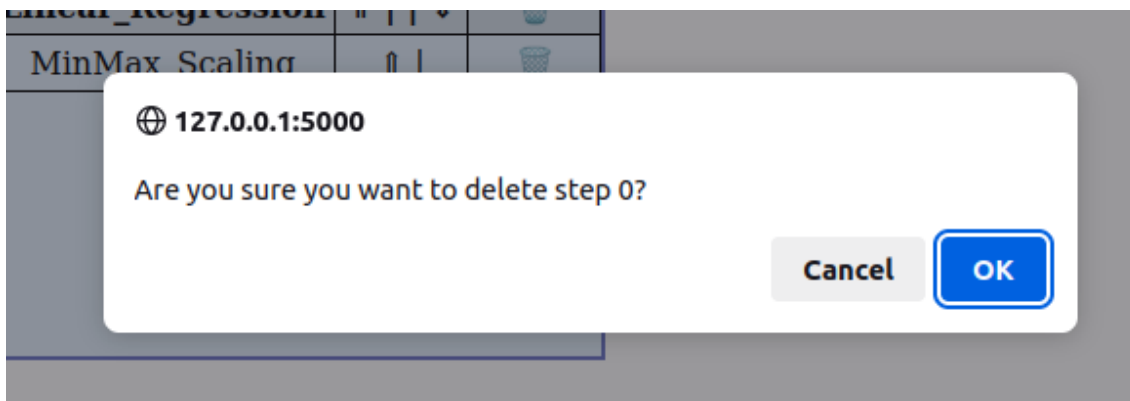
→

Total Steps: 3

Modify enabled

Pos	Steps' Name	Move	Delete
0	MinMax_Scaling	↓	
1	Linear_Regression	↑ ↓	
2	MinMax_Scaling	↑	

2. Eliminar un step: Para ello pulsar el icono de papelera 🗑️. No se puede eliminar el step con el algoritmo principal del pipeline. La operación solo se hará efectiva después de que el usuario confirme la notificación.

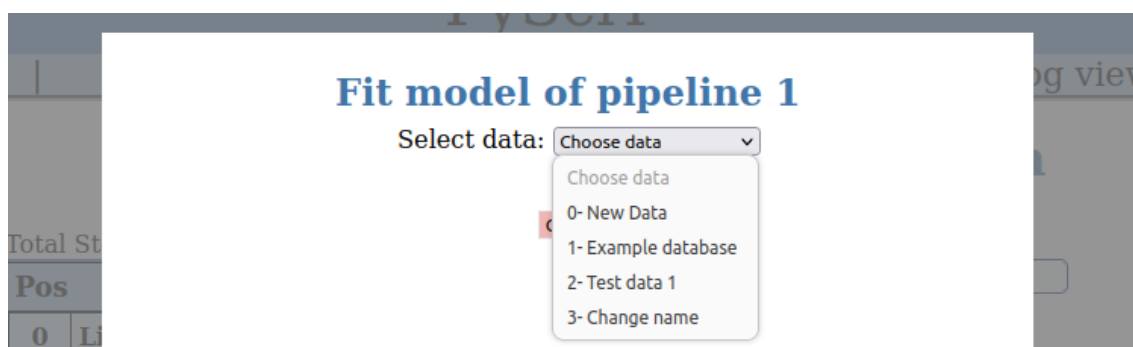


Ejecutar un pipeline

Podremos ejecutar un pipeline pulsando la operación correspondiente en “Pipeline operations”

Pipeline operations	
Fit	Operate
Predict	Operate
Score	Operate

Sea cual sea la operación, se nos abrirá un recuadro para seleccionar una tabla (data) guardada y seleccionar en cada parámetro la columna de datos correspondiente.



Fit model of pipeline 1

Select data: 2- Test data 1

Training data, X: Xtrain1y2

Target values, y: Required

Required

X_train1

X_train2

y_train

Xtrain1y2

Accept

	X_train1	X_train2		Xtrain1y2
0	1	1		[1, 1]
1	1	2	8	[1, 2]
2	2	2	9	[2, 2]
3	2	3	11	[2, 3]

Close

Pulsar “Accept” para ejecutar o “Cancel” para cancelar la operación.

Algunas operaciones tendrán como salida un mensaje que aparecerá en la parte interior de la página. Podremos descargar su contenido con el botón “↑ Save output ↑”.

9/4/2023, 10:53:11 PM- Score

1.0. Used [[[1, 1], [1, 2], [2, 2], [2, 3]], [6, 8, 9, 11]]

↑ Save output ↑

9/4/2023, 10:52:27 PM- Predict

[6. 8. 9. 11.]. Used [[[1, 1], [1, 2], [2, 2], [2, 3]]]

↑ Save output ↑

Close