

Autor: Jefferson Sandoval — Universidad Nacional de Chimborazo Usuario GitHub: deuscheffler

# 1. Introducción

---

## Objetivo del proyecto

El objetivo de este proyecto es desarrollar un flujo completo de preprocesamiento de datos como parte de la asignatura de Ciencia de Datos, implementando técnicas de limpieza, transformación y normalización de conjuntos de datos.

Además, se busca fortalecer el uso de Git y GitHub como herramientas de control de versiones y trabajo colaborativo, incluyendo la automatización mediante GitHub Actions.

## Funcionalidades implementadas

Archivo `preprocesamiento.py`: Implementa funciones para limpiar, codificar y escalar datos.

Control de versiones Git:

Creación del repositorio en GitHub.

Manejo de ramas (feature-preprocesamiento).

Commit y push de los cambios.

Integración y despliegue continuo: Configuración de un flujo automatizado mediante GitHub Actions para verificar el código en cada actualización.

# 2. Comandos Git usados

---

## Comando Descripción

`git clone https://github.com/deuscheffler/preprocesamiento-cienciadatos.git` *Clona el repositorio desde GitHub al entorno local.*

`git config --global user.name "deuscheffler"` *Configura el nombre de usuario en Git.*

`git config --global user.email "jefferson.sandoval@unach.edu.ec"`

```

MINGW64:/c/Users/Jefferson Dell/preprocesamiento-cienciadatos
Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~ (feature-preprocesamiento)
$ git clone https://github.com/deuscheffler/preprocesamiento-cienciadatos.git
Cloning into 'preprocesamiento-cienciadatos'...
remote: Enumerating objects: 7, done.
remote: Counting objects: 100% (7/7), done.
remote: Compressing objects: 100% (6/6), done.
remote: Total 7 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (7/7), 4.16 KiB | 163.00 KiB/s, done.

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~ (feature-preprocesamiento)
$ cd preprocesamiento-cienciadatos

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (main)
$ git config --global user.name "deuscheffler"

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (main)
$ git config --global user.email "jefferson.sandoval@unach.edu.ec"

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (main)
$ git config --list
diff.astextplain.textconv=astextplain
filter.lfs.clean=git-lfs clean -- %f
filter.lfs.smudge=git-lfs smudge -- %f

```

git checkout -b feature-preprocesamiento *Crea y cambia a una nueva rama de desarrollo.*

git add scripts/preprocesamiento.py *Agrega el nuevo archivo al área de preparación.*

git commit -m "Agregar script de preprocesamiento con funciones básicas"

```

MINGW64:/c/Users/Jefferson Dell/preprocesamiento-cienciadatos
credential.https://dev.azure.com.usehttppath=true
init.defaultbranch=master
user.name=deuscheffler
user.email=jefferson.sandoval@unach.edu.ec
user.mail=jefferson.sandoval@unach.edu.ec
core.repositoryformatversion=0
core.filemode=false
core.bare=false
core.logallrefupdates=true
core.symlinks=false
core.ignorecase=true
remote.origin.url=https://github.com/deuscheffler/preprocesamiento-cienciadatos.

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (main)
$ git checkout -b feature-preprocesamiento
Switched to a new branch 'feature-preprocesamiento'

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (feature-preprocesamiento)
$ mkdir scripts

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (feature-preprocesamiento)
$ echo "# Funciones de preprocesamiento de datasets"

```

git checkout main *Cambia a la rama principal.*

git pull origin main *Actualiza la rama principal con los cambios fusionados.*

git branch -d feature-preprocesamiento *Elimina la rama local después de la fusión.*

```
MINGW64:/c/Users/Jefferson Dell/preprocesamiento-cienciadatos - X
Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (feature-preprocesamiento)
$ git add scripts/preprocesamiento.py
warning: in the working copy of 'scripts/preprocesamiento.py', LF will be replaced by CRLF the next time Git touches it

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (feature-preprocesamiento)
$ git commit -m "Aregar script de preprocesamiento con funciones básicas"
[feature-preprocesamiento ba9581a] Agregar script de preprocesamiento con funciones básicas
 1 file changed, 16 insertions(+)
 create mode 100644 scripts/preprocesamiento.py

Jefferson Dell@DESKTOP-VS0C39Q MINGW64 ~/preprocesamiento-cienciadatos (feature-preprocesamiento)
$ git push -u origin feature-preprocesamiento
Enumerating objects: 5, done.
Counting objects: 100% (5/5), done.
Delta compression using up to 4 threads
Compressing objects: 100% (3/3), done.
Writing objects: 100% (4/4), 653 bytes | 50.00 KiB/s, done.
Total 4 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
remote:
remote: Create a pull request for 'feature-preprocesamiento' on GitHub by visiting:
remote:     https://github.com/deuscheffler/preprocesamiento-cienciadatos/pull/
new/feature-preprocesamiento
remote:
To https://github.com/deuscheffler/preprocesamiento-cienciadatos.git
 * [new branch]      feature-preprocesamiento -> feature-preprocesamiento
branch 'feature-preprocesamiento' set up to track 'origin/feature-preprocesamiento'.
```

## Automatización

---

Se creó un flujo automatizado de GitHub Actions para validar el código del proyecto en cada push o pull request. Este flujo se encuentra en el archivo:

.github/workflows/python-app.yml

## All workflows

Filter workflow runs

Showing runs from all workflows

1 workflow run

Event ▾ Status ▾ Branch ▾ Actor ▾

### ✓ Add CI workflow for Python data pr...

CI - Preprocesamiento de Datos #1: Commit

main

ff6a17c pushed by deuscheffler

5 minutes ago ...  
24s

← CI - Preprocesamiento de Datos

### ✓ Add CI workflow for Python data preprocessing #1

Re-run all jobs

...

Summary

Jobs

✓ build

Run details

⌚ Usage

📄 Workflow file

Triggered via push 6 minutes ago

deuscheffler pushed -o ff6a17c main

Status

Success

Total duration

24s

Artifacts

-

python-app.yml

on: push

✓ build

22s

[ ] - +

The screenshot shows a GitHub CI build log for a workflow named "Add CI workflow for Python data preprocessing #1". The build status is "succeeded 6 minutes ago in 22s". The log details the execution of several jobs:

- Set up job (1s)
- Checkout del código (0s)
- Configurar Python (0s)
- Instalar dependencias (17s)
- Validar script de preprocesamiento (0s)
- Post Configurar Python (0s)
- Post Checkout del código (1s)
- Complete job (0s)

On the left sidebar, there are links for "Summary", "Jobs" (selected), "Run details", "Usage", and "Workflow file". A "Re-run all jobs" button is located in the top right corner.

## Pull Request y Fusión en GitHub

Evidencia del proceso de revisión, aprobación y fusión de la rama feature-preprocesamiento con main.



 **preprocesamiento-cienciadatos** Public

 **feature-preprocesamiento** had recent pushes 1 minute ago

[Compare & pull request](#)

 **main**  **2 Branches**  **0 Tags**

Go to file 

 Add file  Code

 **deuscheffler** Rename project and enhance README content 

9ee8ee5 · 1 hour ago  2 Commits

 **.gitignore** Initial commit 1 hour ago

 **README.md** Rename project and enhance README content 1 hour ago

 **README**  

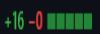
# deuscheffler-preprocesamiento-cienciadatos

## Nombre del proyecto:

Preprocesamiento de Datos - Ciencia de Datos

### Agregar script de preprocesamiento con funciones básicas #1

 **Merged** deuscheffler merged 1 commit into **main** from **feature-preprocesamiento**  9 minutes ago

 Conversation 0  Commits 1  Checks 0  Files changed 1 

 **deuscheffler** commented 9 minutes ago

No description provided.



 **Agregar script de preprocesamiento con funciones básicas**

 Owner 

Reviewers   
No reviews

Still in progress? [Convert to draft](#)

Assignees   
**ba9581a** [assign yourself](#)

Labels   
None yet

 Revert

 **deuscheffler** merged commit **98c3933** into **main** 9 minutes ago

A screenshot of a GitHub pull request merge confirmation screen. At the top, a message from user 'deuscheffler' says 'deleted the feature-preprocesamiento branch now'. Below it, a button labeled 'Restore branch' is visible. To the right, sections for 'Projects' (with 'None yet') and 'Milestone' (with 'No milestone') are shown. A large central box displays a purple icon and the message 'Pull request successfully merged and closed' followed by 'You're all set — the branch has been merged.' Below this is a comment input field with 'Add a comment' placeholder, 'Write' tab selected, and rich text editor icons. A note says 'Successfully merging this pull request may close these issues.' At the bottom, 'None yet' is listed under 'Development'. On the right, 'Notifications' and 'Customize' are shown, with an 'Unsubscribe' button and a note about notifications for modifying open/close state.

A screenshot of a 'Reviewers' modal. It shows a list with one item: 'deuscheffler'. Above the list, it says 'Request up to 15 reviewers'. Below the list, it says 'Nothing to show'. At the bottom, there's a note 'Available on Copilot Pro' with a 'Dismiss' button, and an advertisement for 'Copilot' offering a 30-day free trial.

A screenshot of a GitHub pull request merge confirmation screen. At the top, a message from user 'deuscheffler' says 'restored the feature-preprocesamiento branch now'. Below it, a button labeled 'Delete branch' is visible. To the right, sections for 'Projects' (with 'None yet') and 'Milestone' (with 'No milestone') are shown. A large central box displays a purple icon and the message 'Pull request successfully merged and closed' followed by 'You're all set — the feature-preprocesamiento branch can be safely deleted.' Below this is a comment input field with 'Add a comment' placeholder, 'Write' tab selected, and rich text editor icons.