

# Universidad Nacional de Chimborazo

**Facultad:** Facultad de Ingeniería

**Carrera:** Ciencia de Datos e IA

**Materia:** Cultura digital **Docente:** Ronny Gonzalo Pomboza Granizo

**Semestre:** Tercero

---

Trabajo Autónomo 3

## **Uso de Git y GitHub para la gestión de versiones y la colaboración en proyectos de Ciencia de Datos**

**Autores:**

- Mario Camacho

**Fecha:** 4 de noviembre de 2025

**Riobamba - Ecuador**

---

## Introducción

### Objetivo del Proyecto

Este proyecto busca alicar de una forma practica y eficiente de git combinado con github, que se usara para la gestion de versiones y la colaboracion en diversos proyectos en ciencia de datos

### Funcionalidades Implementadas

Gestion de versiones con git

- creacion y configuracion de un repositorio local y en la nube con github
- Aplicacion de diversas ramas para un mejor desarrollo
- Buen manejo de commits de forma descriptiva y ordenados
- Pull requests para revisiones de codigo y colaboracion del equipo
- Fusion de las ramas para implementar los cambios, ademas de limpieza de ramas obsoletas

### Comandos Git Usados

Configuración Inicial y Clonación

1. Clonar Repositorio

```
git clone https://github.com/tu_usuario/preprocesamiento-cienciadatos.git
```

Lo que hace es crear una copia del repositorio en mi maquina local, lo que ayuda a crear un entorno de trabajo que este sincronizado con github para comenzar un desarrollo

## 2. Configurar Usuario y Correo

```
git config --global user.name "Tu Nombre"  
git config --global user.email "tuemail@example.com"
```

Define la identidad del usuario del que esta trabajando y se le asignara a los commits, lo que ayuda a identificar quien hizo los cambios en el proyecto en trabajos colaborativos

## Gestión de Ramas y Flujo de Trabajo

### 3. Crear Nueva Rama

```
git checkout -b feature-preprocesamiento
```

Lo que hace es crear una nueva rama a partir de la principal, e inmediatamente cambia a ella. lo que ayuda a desarrollar en un entorno aislado para agregar nuevas funcionalidades sin afectar la rama principal

## Control de Cambios y Versionado

### 4. Agregar Cambios al Área de Preparación

```
git add .
```

Prepara todos los ficheros que han sido modificados con el objetivo de incluirlos en el proximo commit, lo que permite seleccionar los cambios que se versionar

### 5. Hacer un Commit

```
git commit -m "Aregar función para preprocesamiento completo de datos"
```

Guarda todos los cambios realizados con un mensaje informativo, lo que ayuda a tener un historial claro y en colaboraciones ayuda a entender el trabajo de los demás

## Sincronización con Repositorio Remoto

### 6. Subir Cambios a GitHub

```
git push origin feature-preprocesamiento
```

Envia todos los cambios guardados de forma local al repositorio remoto, lo que permite compartir todo el trabajo modificado con un equipo ademas de guardar una copia de seguridad en la nube

#### 7. Actualizar Repositorio Local

```
git pull origin main
```

Obtiene los últimos cambios del repositorio remoto y los fusiona con la rama local actual, lo que mantiene el desarrollo de la rama lo mas actualizado posible con los cambios realizados por otros colaboradores

#### 8. Fusionar Ramas

```
git checkout main  
git merge feature-preprocesamiento
```

integra todos los cambios realizados en una rama a la rama principal, lo que permite consolidar el trabajo realizado y avanzar en el desarrollo del proyecto

### Limpieza y Mantenimiento

#### 9. Eliminar Rama Local

```
git branch -d feature-preprocesamiento
```

Elimina la rama que se uso para un desarrollo, lo que ayuda a mantener todo el reppositorio limpio y organizado, permite quitar ramas con desarrollo obsoleto

## Automatización: Explicación del Workflow de GitHub Actions Creado

Se implemento un workflow que esta integrado en github actions y lo que hace es automatizas la verificacion del preprocesamiento de los datos cada vez que se realice un oush o algun cambio en el codigo

### Funcionamiento del Workflow

#### Trigger de Ejecución

- Eventos activadores: push y pull\_request

este se ejecuta cada ves que detecta algun tip ode cambio en el codigo que se sube en el repositorio

#### Proceso Automatizado

- Entorno limpio: Se crea máquina virtual Ubuntu desde cero
- Preparación:

- Descarga del código actual
- Instalación de Python 3.8
- Instalación de librerías (pandas, scikit-learn, numpy, openpyxl)
- Ejecución:
  - Crea estructura de carpetas necesaria
  - Ejecuta el script completo de preprocesamiento
- Verificación:
  - Confirma que se generaron los archivos esperados
  - Reporta éxito o fallo automáticamente

## Resultados Esperados

- Build exitoso: Todos los pasos se completan sin errores
- Archivos generados: data/dataset\_ejemplo.csv y outputs/dataset\_procesado.csv
- Logs detallados: Output del preprocesamiento visible en GitHub

## Capturas de Pantalla

```
PS C:\Users\mario\OneDrive\Desktop\trabajo> git clone https://github.com/RosariosTijeras/preprocesamiento-cienciadatos.git
Cloning into 'preprocesamiento-cienciadatos'...
remote: Enumerating objects: 16, done.
remote: Counting objects: 100% (16/16), done.
remote: Compressing objects: 100% (15/15), done.
remote: Total 16 (delta 3), reused 0 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (16/16), 6.89 KiB | 1.38 MiB/s, done.
Resolving deltas: 100% (3/3), done.
PS C:\Users\mario\OneDrive\Desktop\trabajo> git config --global user.name "RosariosTijeras"
PS C:\Users\mario\OneDrive\Desktop\trabajo> git config --global user.email "mario.camacho@unach.edu.ec"
PS C:\Users\mario\OneDrive\Desktop\trabajo> cd ./preprocesamiento-cienciadatos\
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos> git checkout -b feature-preprocesamiento
Switched to a new branch 'feature-preprocesamiento'

PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos\src> git add preprocesamiento.py
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos\src> git commit -m "Creacion de un archivo con funciones completas para el preprocesamiento de un dataset"
[feature-preprocesamiento db95e91] Creacion de un archivo con funciones completas para el preprocesamiento de un dataset
 1 file changed, 0 insertions(+), 0 deletions(-)
 create mode 100644 src/preprocesamiento.py
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos\src> git push origin feature-preprocesamiento
Enumerating objects: 5, done.
Counting objects: 100% (5/5), done.
Delta compression using up to 12 threads
Compressing objects: 100% (2/2), done.
Writing objects: 100% (4/4), 421 bytes | 421.00 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/RosariosTijeras/preprocesamiento-cienciadatos/pull/new/feature-preprocesamiento
remote:
To https://github.com/RosariosTijeras/preprocesamiento-cienciadatos.git
 * [new branch]      feature-preprocesamiento -> feature-preprocesamiento
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos\src> |
```

A screenshot of a GitHub repository interface. At the top, a notification bar says "feature-preprocesamiento had recent pushes on 4 nov" with a "Compare & pull request" button. Below it, a navigation bar shows "main" (selected), "1 Branch", "0 Tags", a search bar ("Go to file"), and buttons for "Add file" and "Code". A sidebar on the right has icons for "Issues", "Pull requests", "Commits", "File browser", "Repository info", and "Settings". The main content area shows a commit from "RosariosTijeras" titled "arreglando readme.md" made 10 minutes ago. It includes files ".gitignore" (initial commit, 1 hour ago) and "README.md" (arreglando readme.md, 10 minutes ago). A "README" file is currently selected.

## Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also compare across forks. [Learn more about diff commits](#)

A screenshot of the GitHub pull request creation interface. At the top, it shows "base: main" and "compare: feature-preprocesamiento" with a green checkmark indicating "Able to merge". Below that, there's a "Add a title" field with placeholder text "Creacion de un archivo con funciones completas para el preprocesamien..." and a rich text editor toolbar. The editor has tabs for "Write" and "Preview", and various formatting buttons like bold, italic, and code. In the preview area, the text "...to de un dataset" is shown. At the bottom left, it says "Markdown is supported" and "Paste, drop, or click to add files". At the bottom right, there's a large green "Create pull request" button.

RosariosTijeras commented now

...to de un dataset

Mention @copilot in a comment to make changes to this pull request.

Creation of a file with complete functions for the preprocessing...

Commit message

Merge pull request #1 from RosariosTijeras/feature-preprocesamiento

Extended description

Creation of a file with complete functions for the preprocessing...

This commit will be authored by mario.camacho@unach.edu.ec.

Confirm merge Cancel

preprocesamiento-cienciadatos Public

main 1 Branch 0 Tags

Go to file Add file Code

RosariosTijeras Merge pull request #1 from RosariosTijeras/feature-preprocesamiento 30ba599 · 4 minutes ago 7 Commits

src Creation of a file with complete functions for the pre... 6 minutes ago

.gitignore Initial commit 1 hour ago

README.md arregloando readme.md 16 minutes ago

README

# Proyecto de Preprocesamiento en Ciencia de Datos

## Objetivo

```
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos> git add .github/workflows/test.yml
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos> git add src/preprocesamiento.py
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos> git commit -m "Aregar GitHub Actions para automatización de pruebas"
[feature-preprocesamiento 32a1562] Agregar GitHub Actions para automatización de pruebas
 2 files changed, 239 insertions(+)
  create mode 100644 .github/workflows/test.yml
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos> git push origin feature-preprocesamiento
Enumerating objects: 10, done.
Counting objects: 100% (10/10), done.
Delta compression using up to 12 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (7/7), 3.12 KiB | 1.56 MiB/s, done.
Total 7 (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/RosariosTijeras/preprocesamiento-cienciadatos/pull/new/feature-preprocesamiento
remote:
To https://github.com/RosariosTijeras/preprocesamiento-cienciadatos.git
 * [new branch]      feature-preprocesamiento -> feature-preprocesamiento
PS C:\Users\mario\OneDrive\Desktop\trabajo\preprocesamiento-cienciadatos>
```

## Repositorio del Proyecto

<https://github.com/RosariosTijeras/preprocesamiento-cienciadatos.git>

## Bibliografía

- Chacon, S., & Straub, B. (2014). *Pro Git* (2nd ed.). Apress. <https://git-scm.com/book/en/v2>
- GitHub. (n.d.). *GitHub Docs: Get started using Git*. <https://docs.github.com/en/get-started/using-git>
- Chacon, S., & Straub, B. (2014). *Pro Git*. Apress
- McKinney, W. (2018). *Python for Data Analysis*. O'Reilly Media.