

Manual de Usuario para la aplicación pgWatuyML

Universidad de Nariño, Colombia

Versión 1.0

GNU General Public License

2025



Indice

1. GENERALIDADES DEL SISTEMA.....	3
1.1. Objetivo.....	3
1.2. Descarga del software.....	3
1.3. Licencia.....	3
2. INSTALACIÓN DE LA APLICACIÓN.....	4
2.1. Instalar PostgreSQL y dependencias para las extensiones.....	4
2.2. Instalar extensiones para PostgreSQL.....	6
2.3. Instalar herramienta pgWatuyML.....	11
2.3.1. Instalar desde archivo (Ubuntu).....	11
2.3.2. Compilar la herramienta pgWatuyML.....	12
3. SECCIONES DE LA APLICACIÓN.....	19
3.1. Conexión.....	19
3.2. Preprocesamiento.....	20
3.2.1. Particionar Tabla.....	20
3.2.2. Cortar tabla.....	22
3.2.3. Importar desde CSV.....	23
3.3. Clasificación.....	24
3.4. Clustering.....	26

1. GENERALIDADES DEL SISTEMA

1.1. Objetivo

pgWatuyML es una herramienta gráfica de minería de datos para PostgreSQL, diseñada especialmente para usuarios no técnicos. Su interfaz intuitiva permite realizar diversas tareas de preprocesamiento, tales como la división de tablas por porcentaje, la importación de archivos CSV a las tablas y la creación de nuevas tablas a partir de columnas seleccionadas. Además, integra algoritmos de clasificación, como los árboles de decisión C4.5 e ID3, así como algoritmos de clustering, entre los que destaca K-Means.

1.2. Descarga del software

Para descargar el archivo comprimido con el código fuente, instaladores y documentación puede hacerlo a través del sitio web de GRIAS mediante el siguiente enlace <http://grias.udenar.edu.co/grias/?p=598>.

1.3. Licencia

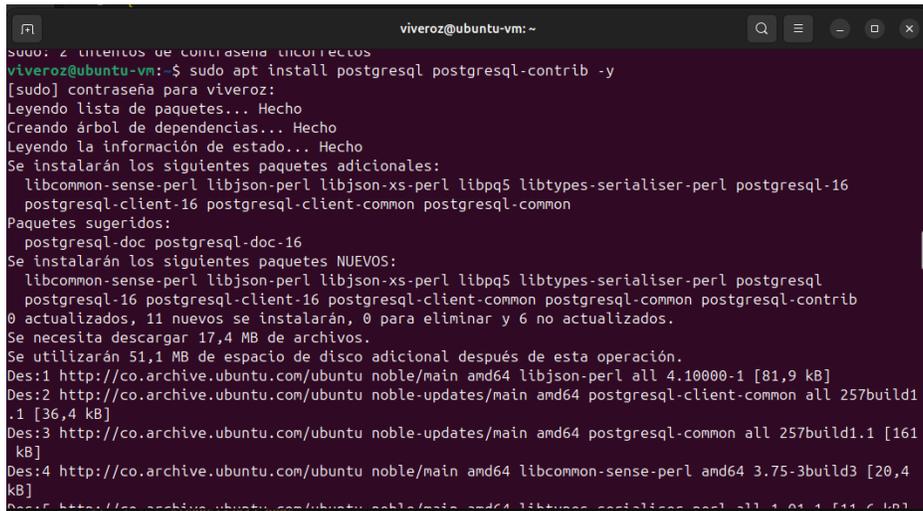
pgWatuyML se distribuye bajo la licencia GPL v3 (GNU General Public License versión 3), lo que garantiza que el software es libre y de código abierto. Esto permite a los usuarios utilizarlo, modificarlo y redistribuirlo, siempre que se mantenga la misma licencia en las versiones derivadas.

2. INSTALACIÓN DE LA APLICACIÓN

2.1. Instalar PostgreSQL y dependencias para las extensiones

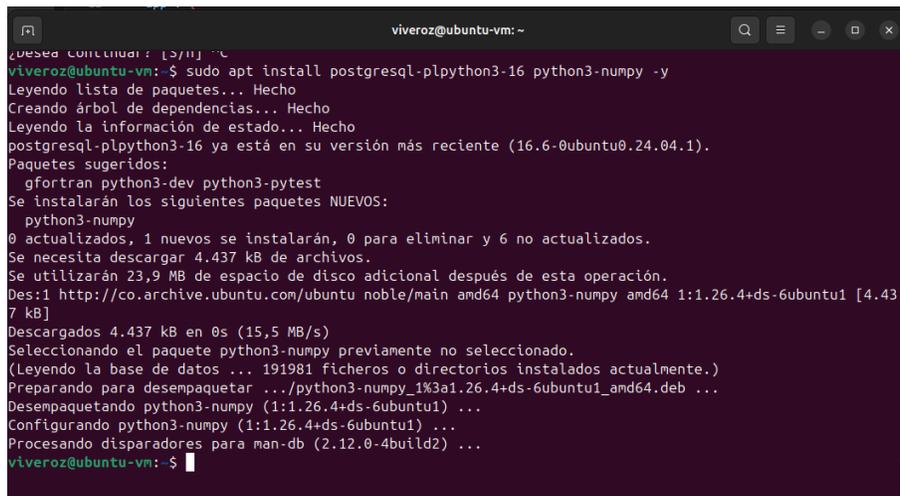
Es necesario instalar la base de datos PostgreSQL para

- 1) Abrir una terminal con **CTRL + ALT + T** y escribir el siguiente comando para instalar la base de datos PostgreSQL: ***sudo apt install postgresql postgresql-contrib -y***



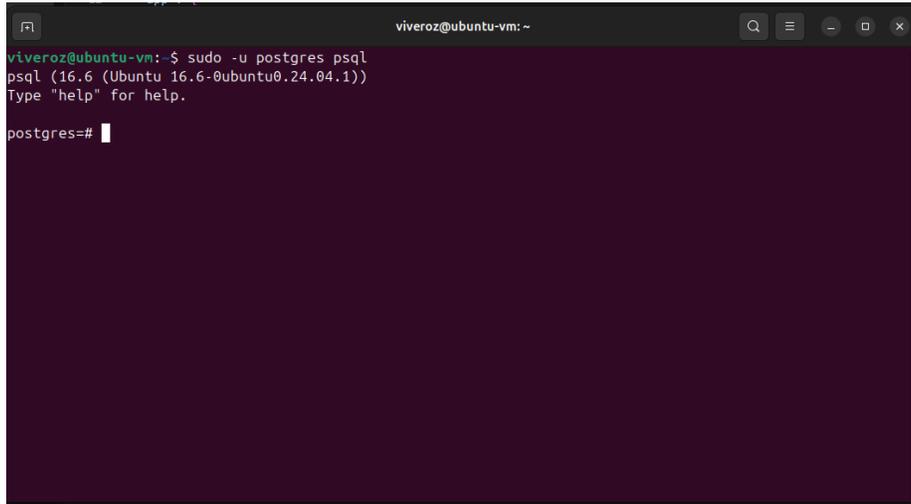
```
viveroz@ubuntu-vm: ~$ sudo apt install postgresql postgresql-contrib -y
[sudo] contraseña para viveroz:
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Se instalarán los siguientes paquetes adicionales:
 libcommon-sense-perl libjson-perl libjson-xs-perl libpq5 libtypes-serialiser-perl postgresql-16
 postgresql-client-16 postgresql-client-common postgresql-common
Paquetes sugeridos:
 postgresql-doc postgresql-doc-16
Se instalarán los siguientes paquetes NUEVOS:
 libcommon-sense-perl libjson-perl libjson-xs-perl libpq5 libtypes-serialiser-perl postgresql
 postgresql-16 postgresql-client-16 postgresql-client-common postgresql-common postgresql-contrib
0 actualizados, 11 nuevos se instalarán, 0 para eliminar y 6 no actualizados.
Se necesita descargar 17,4 MB de archivos.
Se utilizarán 51,1 MB de espacio de disco adicional después de esta operación.
Des:1 http://co.archive.ubuntu.com/ubuntu noble/main amd64 libjson-perl all 4.10000-1 [81,9 kB]
Des:2 http://co.archive.ubuntu.com/ubuntu noble-updates/main amd64 postgresql-client-common all 257build1
.1 [36,4 kB]
Des:3 http://co.archive.ubuntu.com/ubuntu noble-updates/main amd64 postgresql-common all 257build1.1 [161
kB]
Des:4 http://co.archive.ubuntu.com/ubuntu noble/main amd64 libcommon-sense-perl amd64 3.75-3build3 [20,4
kB]
Des:5 http://co.archive.ubuntu.com/ubuntu noble/main amd64 libtypes-serialiser-perl all 4.01-1 [11,6 kB]
```

- 2) Para el correcto funcionamiento de las extensiones es necesario instalar el lenguaje procedimental **plpython**, al igual que la librería de python **numpy**, para lo cual se ejecuta el siguiente comando: ***sudo apt install postgresql-plpython3-16 python3-numpy -y***



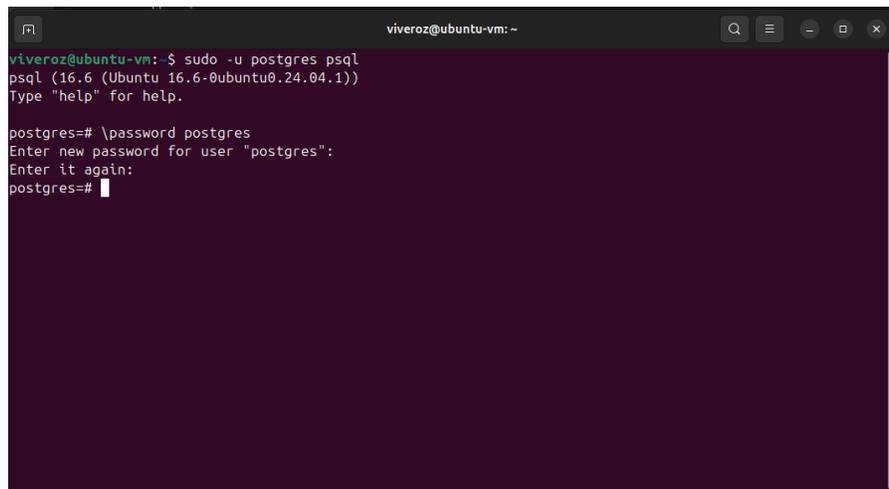
```
viveroz@ubuntu-vm: ~$ sudo apt install postgresql-plpython3-16 python3-numpy -y
¿desea continuar? [y/n] y
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
postgresql-plpython3-16 ya está en su versión más reciente (16.6-0ubuntu0.24.04.1).
Paquetes sugeridos:
  gfortran python3-dev python3-pytest
Se instalarán los siguientes paquetes NUEVOS:
  python3-numpy
0 actualizados, 1 nuevos se instalarán, 0 para eliminar y 6 no actualizados.
Se necesita descargar 4.437 kB de archivos.
Se utilizarán 23,9 MB de espacio de disco adicional después de esta operación.
Des:1 http://co.archive.ubuntu.com/ubuntu noble/main amd64 python3-numpy amd64 1:1.26.4+ds-6ubuntu1 [4.43
7 kB]
Descargados 4.437 kB en 0s (15,5 MB/s)
Seleccionando el paquete python3-numpy previamente no seleccionado.
(Leyendo la base de datos ... 191981 ficheros o directorios instalados actualmente.)
Preparando para desempaquetar .../python3-numpy_1%3a1.26.4+ds-6ubuntu1_amd64.deb ...
Desempaquetando python3-numpy (1:1.26.4+ds-6ubuntu1) ...
Configurando python3-numpy (1:1.26.4+ds-6ubuntu1) ...
Procesando disparadores para man-db (2.12.0-4build2) ...
viveroz@ubuntu-vm: ~$
```

- 3) Si el primera vez que se instala **PostgreSQL** es necesario crear una contraseña para el usuario **postgres**, y opcionalmente una base de datos, para interactuar con la base de datos utilice **psql**, mediante el siguiente comando: **sudo -u postgres psql**



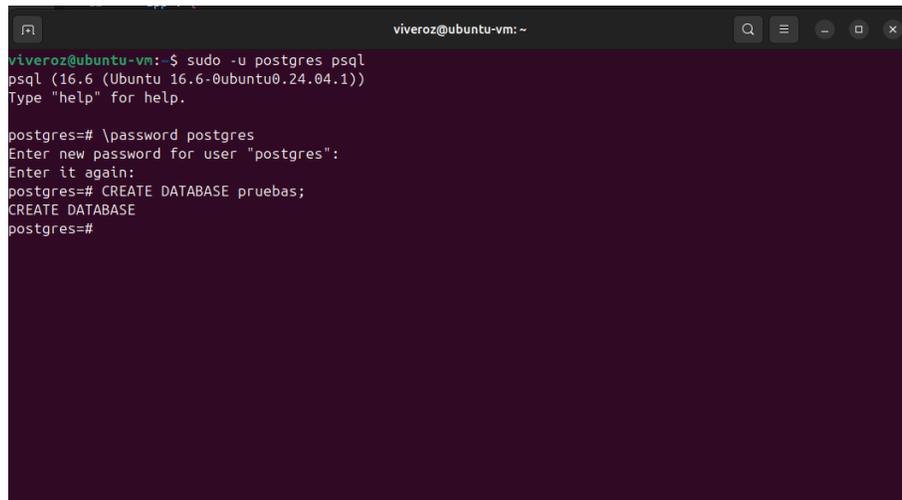
```
viveroz@ubuntu-vm: ~  
viveroz@ubuntu-vm:~$ sudo -u postgres psql  
psql (16.6 (Ubuntu 16.6-0ubuntu0.24.04.1))  
Type "help" for help.  
  
postgres=#
```

- 4) Para cambiar la contraseña del usuario **postgres** escriba el siguiente comando y pulse ENTER, (el comando le pedirá ingresar la nueva contraseña 2 veces): **\password postgres**



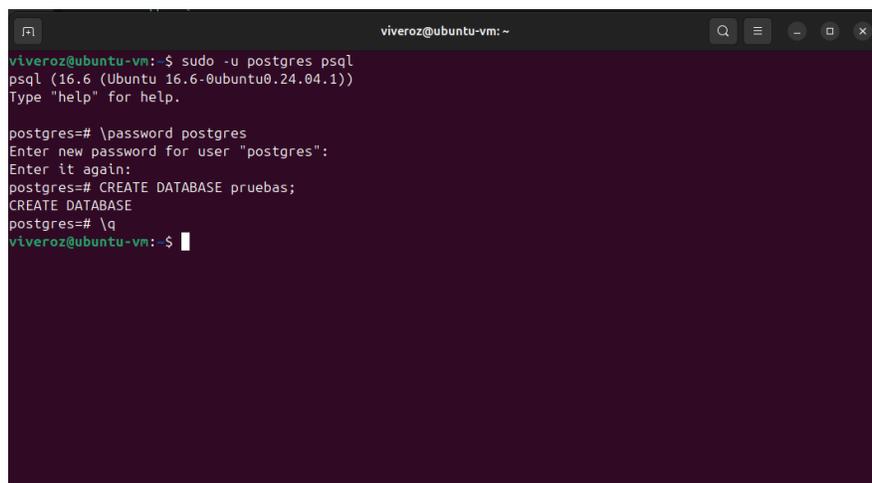
```
viveroz@ubuntu-vm: ~  
viveroz@ubuntu-vm:~$ sudo -u postgres psql  
psql (16.6 (Ubuntu 16.6-0ubuntu0.24.04.1))  
Type "help" for help.  
  
postgres=# \password postgres  
Enter new password for user "postgres":  
Enter it again:  
postgres=#
```

- 5) Puede crear una base de datos con el comando **CREATE DATABASE <nombre>**; por ejemplo **CREATE DATABASE pruebas**;



```
viveroz@ubuntu-vm: ~  
viveroz@ubuntu-vm:~$ sudo -u postgres psql  
psql (16.6 (Ubuntu 16.6-0ubuntu0.24.04.1))  
Type "help" for help.  
  
postgres=# \password postgres  
Enter new password for user "postgres":  
Enter it again:  
postgres=# CREATE DATABASE pruebas;  
CREATE DATABASE  
postgres=#
```

- 6) Puede salir de psql escribiendo el comando **\q**



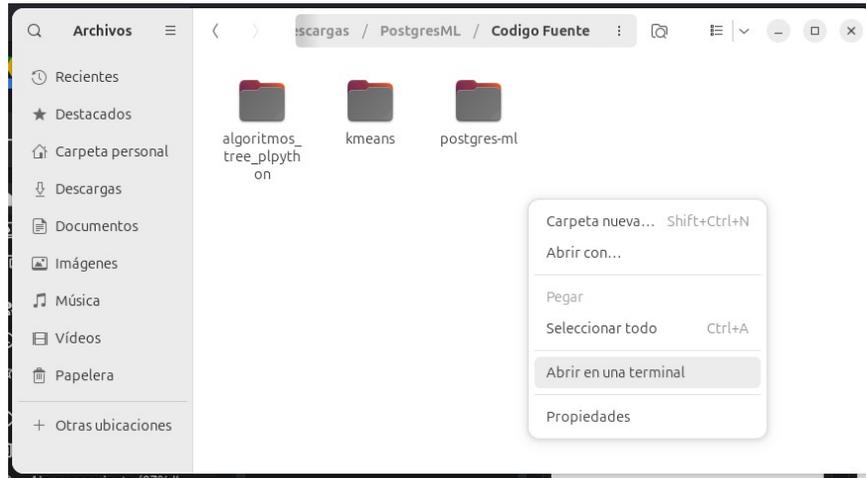
```
viveroz@ubuntu-vm: ~  
viveroz@ubuntu-vm:~$ sudo -u postgres psql  
psql (16.6 (Ubuntu 16.6-0ubuntu0.24.04.1))  
Type "help" for help.  
  
postgres=# \password postgres  
Enter new password for user "postgres":  
Enter it again:  
postgres=# CREATE DATABASE pruebas;  
CREATE DATABASE  
postgres=# \q  
viveroz@ubuntu-vm:~$
```

7. Puede cerrar la terminal si lo desea.

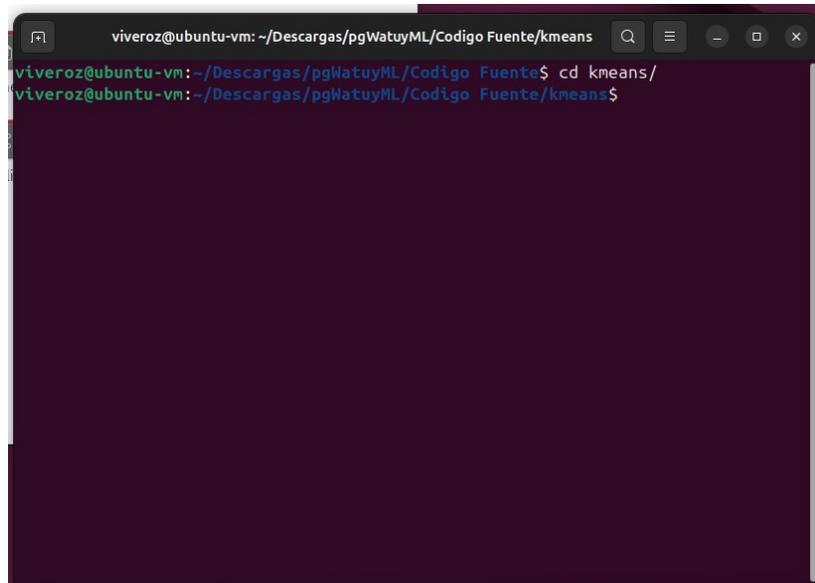
2.2. Instalar extensiones para PostgreSQL

Para que la herramienta funcione de manera adecuada es necesario instalar 2 extensiones para PostgreSQL

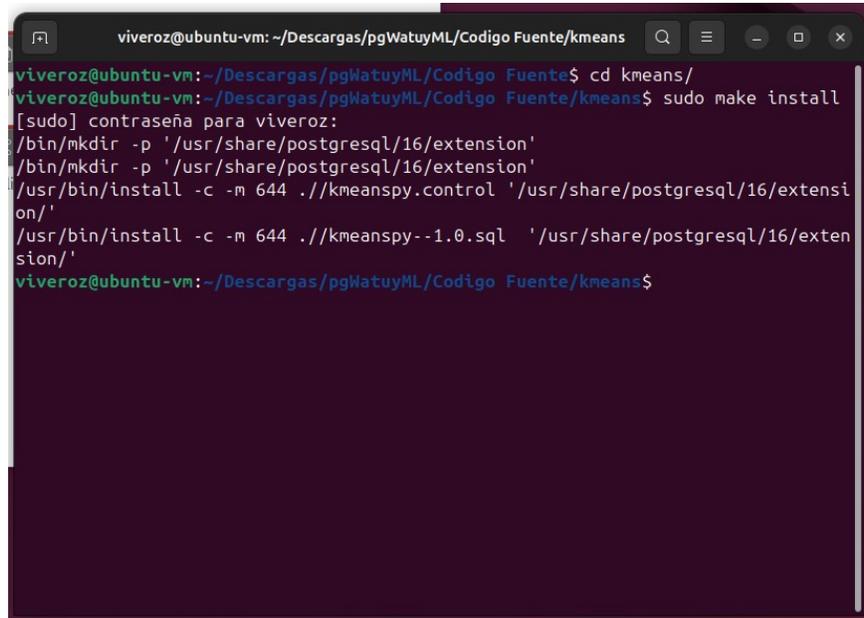
- 1) Abra el explorador de archivos, diríjase a la carpeta donde descomprimió el proyecto, luego entre a **Código Fuente** y después haga clic derecho en el espacio vacío en la carpeta y pulse en **Abrir en una terminal**, esto abrirá una terminal en el directorio actual



- 2) Escriba el siguiente comando para cambiar el directorio de trabajo a kmeans: **cd kmeans**

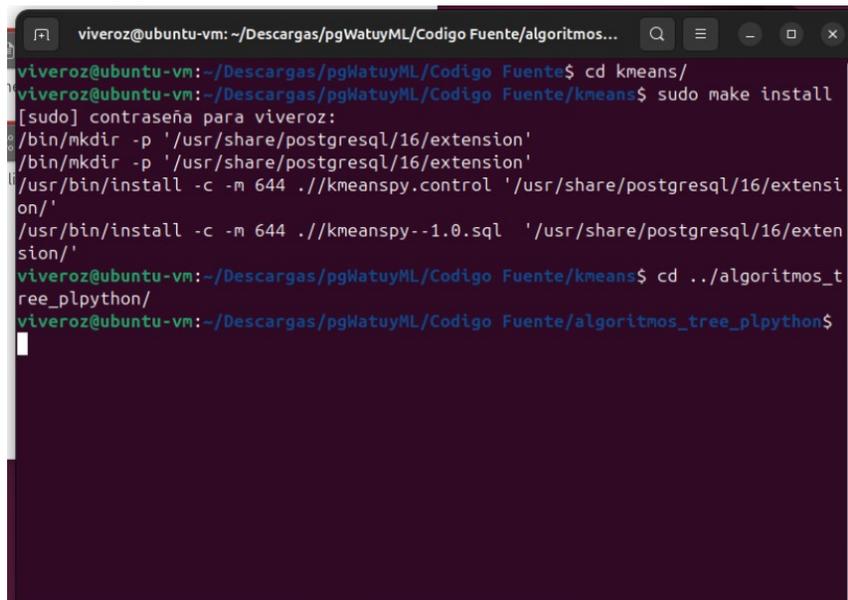
A screenshot of a terminal window. The title bar reads 'viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/kmeans'. The terminal shows the prompt 'viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente\$' followed by the command 'cd kmeans/' and the resulting prompt 'viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/kmeans\$'.

3) Para instalar la extensión en PostgreSQL escriba el comando: `sudo make install`



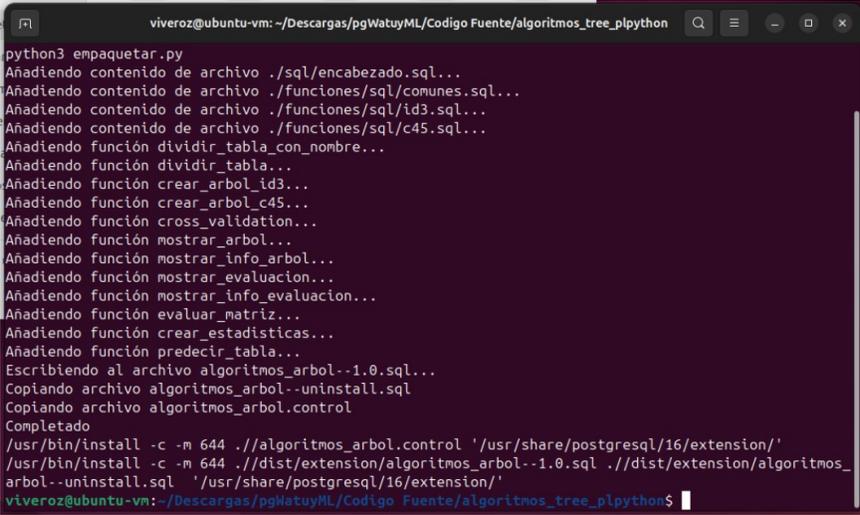
```
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/kmeans
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente$ cd kmeans/
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/kmeans$ sudo make install
[sudo] contraseña para viveroz:
/bin/mkdir -p '/usr/share/postgresql/16/extension'
/bin/mkdir -p '/usr/share/postgresql/16/extension'
/usr/bin/install -c -m 644 ./kmeanspy.control '/usr/share/postgresql/16/extension/'
/usr/bin/install -c -m 644 ./kmeanspy--1.0.sql '/usr/share/postgresql/16/extension/'
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/kmeans$
```

4) Escriba el siguiente comando para navegar hasta la carpeta con la otra extensión: `cd ../algoritmos_tree_plpython/`



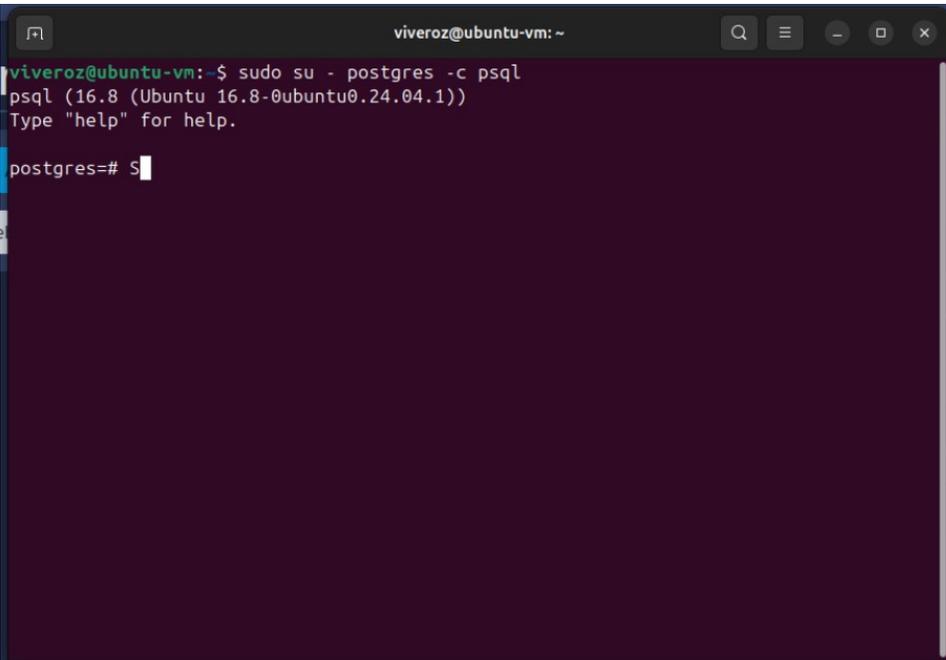
```
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/algoritmos...
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/kmeans$ sudo make install
[sudo] contraseña para viveroz:
/bin/mkdir -p '/usr/share/postgresql/16/extension'
/bin/mkdir -p '/usr/share/postgresql/16/extension'
/usr/bin/install -c -m 644 ./kmeanspy.control '/usr/share/postgresql/16/extension/'
/usr/bin/install -c -m 644 ./kmeanspy--1.0.sql '/usr/share/postgresql/16/extension/'
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/kmeans$ cd ../algoritmos_tree_plpython/
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/algoritmos_tree_plpython$
```

5) Escriba de nuevo el comando para instalar la extensión: **sudo make install**



```
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/algoritmos_tree_plpython
python3 empaquetar.py
Añadiendo contenido de archivo ./sql/encabezado.sql...
Añadiendo contenido de archivo ./funciones/sql/comunes.sql...
Añadiendo contenido de archivo ./funciones/sql/id3.sql...
Añadiendo contenido de archivo ./funciones/sql/c45.sql...
Añadiendo función dividir_tabla_con_nombre...
Añadiendo función dividir_tabla...
Añadiendo función crear_arbol_id3...
Añadiendo función crear_arbol_c45...
Añadiendo función cross_validation...
Añadiendo función mostrar_arbol...
Añadiendo función mostrar_info_arbol...
Añadiendo función mostrar_evaluacion...
Añadiendo función mostrar_info_evaluacion...
Añadiendo función evaluar_matriz...
Añadiendo función crear_estadisticas...
Añadiendo función predecir_tabla...
Escribiendo al archivo algoritmos_arbol--1.0.sql...
Copiando archivo algoritmos_arbol--uninstall.sql
Copiando archivo algoritmos_arbol.control
Completado
/usr/bin/install -c -m 644 ./algoritmos_arbol.control '/usr/share/postgresql/16/extension/'
/usr/bin/install -c -m 644 ./dist/extension/algoritmos_arbol--1.0.sql ./dist/extension/algoritmos_arbol--uninstall.sql '/usr/share/postgresql/16/extension/'
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/algoritmos_tree_plpython$
```

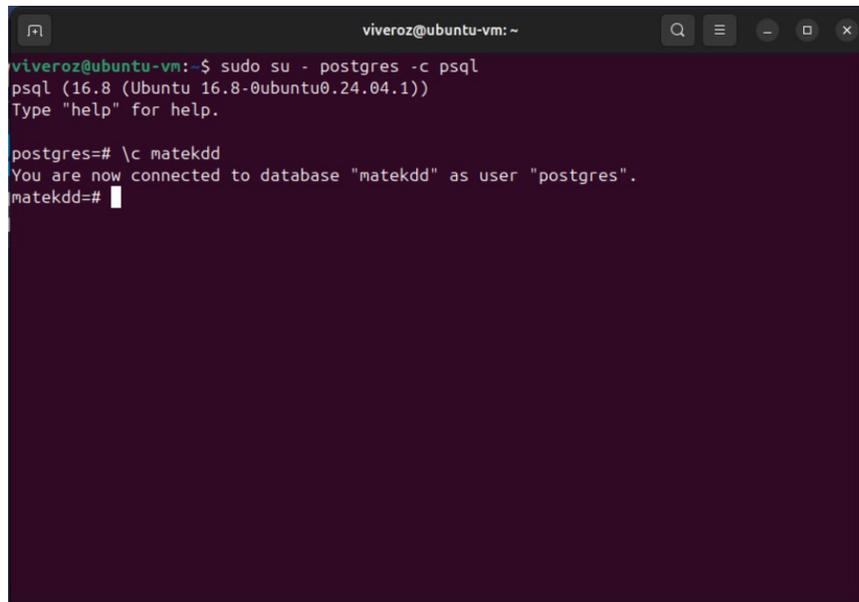
6) Ahora ejecute la consola de PostgreSQL con el usuario postgres mediante el comando: **sudo su - postgres -c psql**



```
viveroz@ubuntu-vm: ~
viveroz@ubuntu-vm:~$ sudo su - postgres -c psql
psql (16.8 (Ubuntu 16.8-0ubuntu0.24.04.1))
Type "help" for help.

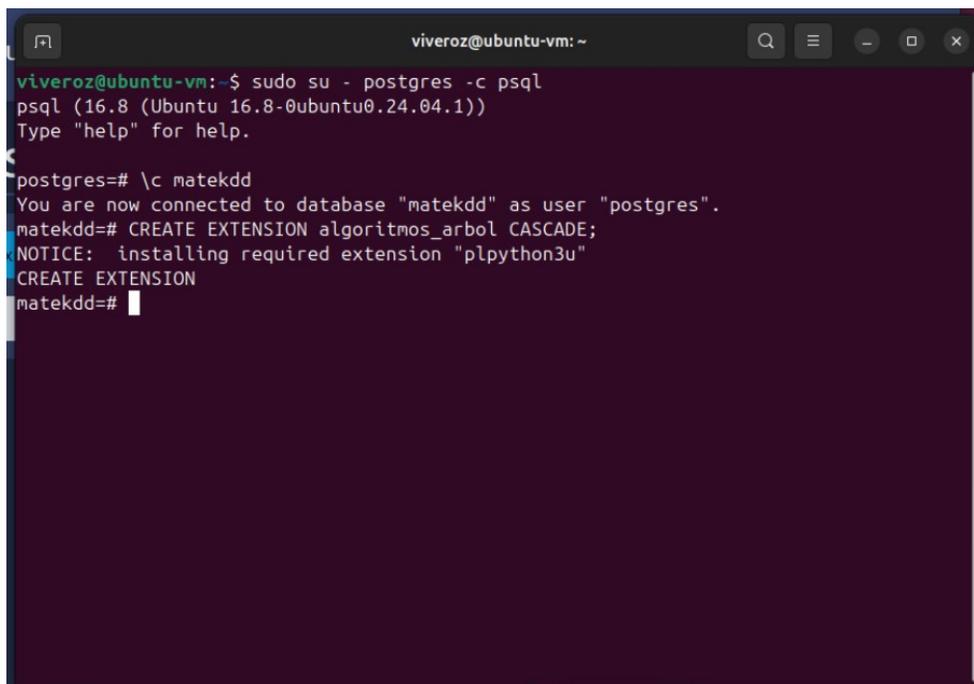
postgres=# S
```

7) Se puede conectar una base de datos con: \c <base_de_datos>

A terminal window titled 'viveroz@ubuntu-vm: ~' showing a sequence of commands to connect to a PostgreSQL database. The user runs 'sudo su - postgres -c psql', which opens a psql prompt. Then, they run '\c matekdd', which connects them to the 'matekdd' database as the 'postgres' user. The prompt changes from 'psql (16.8 (Ubuntu 16.8-0ubuntu0.24.04.1))' to 'matekdd=#'.

```
viveroz@ubuntu-vm: ~  
viveroz@ubuntu-vm:~$ sudo su - postgres -c psql  
psql (16.8 (Ubuntu 16.8-0ubuntu0.24.04.1))  
Type "help" for help.  
  
postgres=# \c matekdd  
You are now connected to database "matekdd" as user "postgres".  
matekdd=#
```

8) Para instalar la extensión de algoritmos de árbol ingrese el comando: **CREATE EXTENSION algoritmos_arbol CASCADE;**

A terminal window titled 'viveroz@ubuntu-vm: ~' showing the installation of a PostgreSQL extension. The user runs 'CREATE EXTENSION algoritmos_arbol CASCADE;' in the psql prompt. The system responds with a notice: 'NOTICE: installing required extension "plpython3u"'. The prompt then returns to 'matekdd=#'.

```
viveroz@ubuntu-vm: ~  
viveroz@ubuntu-vm:~$ sudo su - postgres -c psql  
psql (16.8 (Ubuntu 16.8-0ubuntu0.24.04.1))  
Type "help" for help.  
  
postgres=# \c matekdd  
You are now connected to database "matekdd" as user "postgres".  
matekdd=# CREATE EXTENSION algoritmos_arbol CASCADE;  
NOTICE: installing required extension "plpython3u"  
CREATE EXTENSION  
matekdd=#
```

- 9) Para instalar la extensión de K-means ingrese el comando: **CREATE EXTENSION kmeanspy CASCADE;**

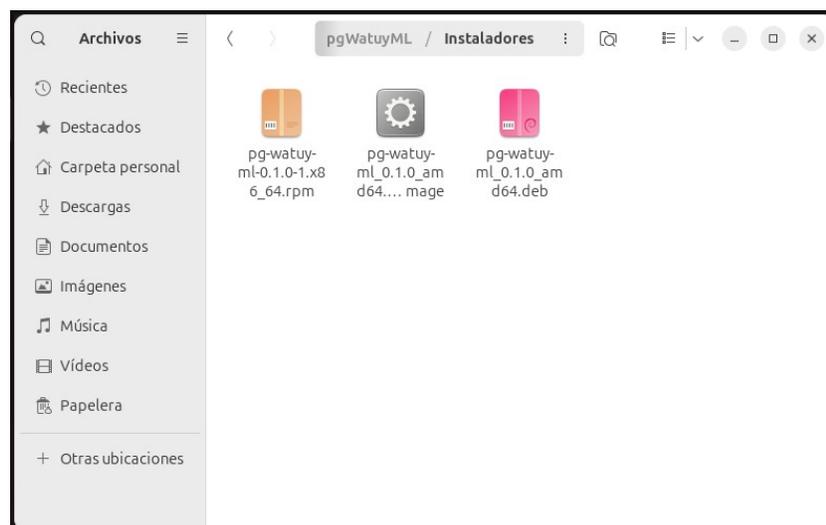
```
viveroz@ubuntu-vm: ~  
viveroz@ubuntu-vm:~$ sudo su - postgres -c psql  
psql (16.8 (Ubuntu 16.8-0ubuntu0.24.04.1))  
Type "help" for help.  
  
postgres=# \c matekdd  
You are now connected to database "matekdd" as user "postgres".  
matekdd=# CREATE EXTENSION algoritmos_arbol CASCADE;  
NOTICE: installing required extension "plpython3u"  
CREATE EXTENSION  
matekdd=# CREATE EXTENSION kmeanspy CASCADE;  
CREATE EXTENSION  
matekdd=#
```

- 10) Puede cerrar la terminal si lo desea.

2.3. Instalar herramienta pgWatuyML

2.3.1. Instalar desde archivo (Ubuntu)

- 1) Abra el explorador de archivos, diríjase a la carpeta donde descomprimió el proyecto, luego entre a **Instaladores** y después haga clic doble click sobre el archivo .deb para instalar la aplicación.



2) Haga clic en **Instalar**.



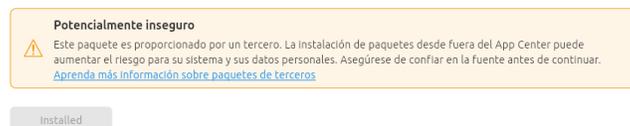
3) Aparecerá una ventana solicitando confirmación para instalar, haga clic en **Instalar**.



4) Ingrese su contraseña y haga clic en **Autenticar**



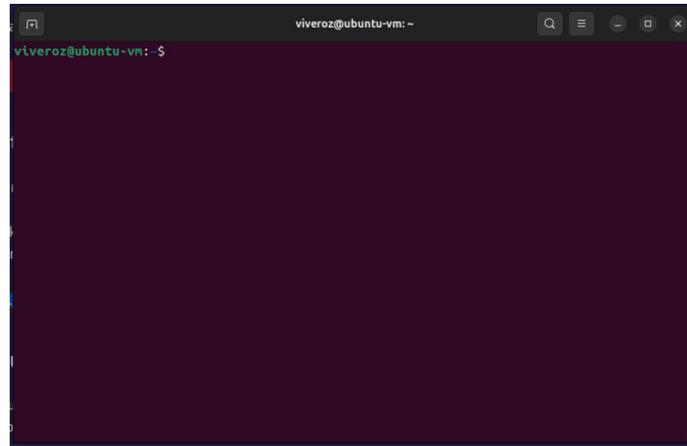
5) La aplicación debería instalarse de manera exitosa mostrando el texto **Installed**.



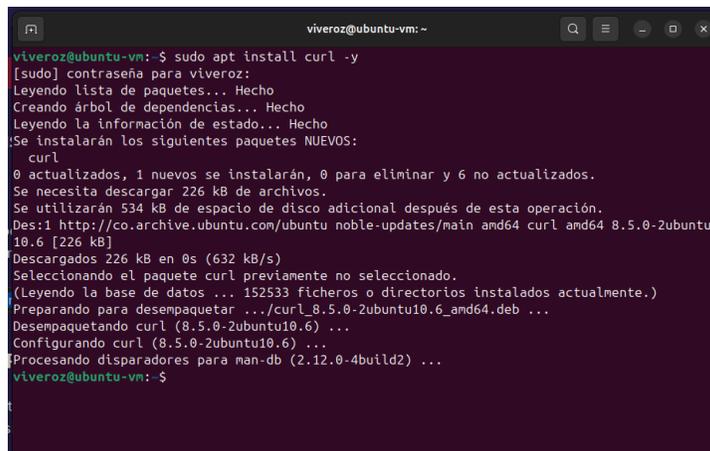
2.3.2. **Compilar la herramienta pgWatuyML**

En caso de que requiera modificar el código fuente en la aplicación y compilar una nueva versión puede seguir los pasos descritos en esta sección.

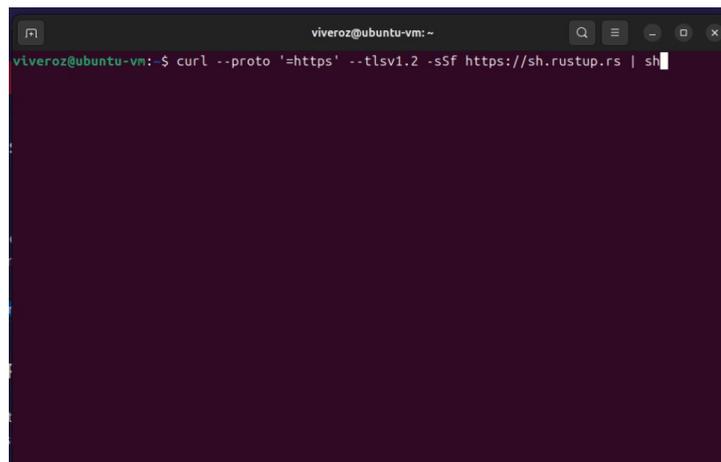
1) Abra una terminal pulsando la combinación de teclas CTRL + ALT + T



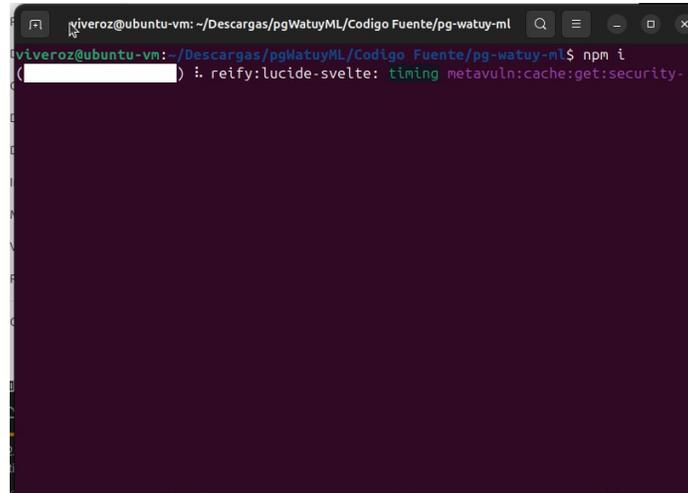
2) Es necesario instalar **Curl**, puede hacerlo con el comando: **sudo apt install curl -y**



3) Para compilar el programa se requiere tener **Rust** instalado, lo puede instalar con el comando: **curl --proto '=https' --tlsv1.2 -sSf https://sh.rustup.rs | sh**

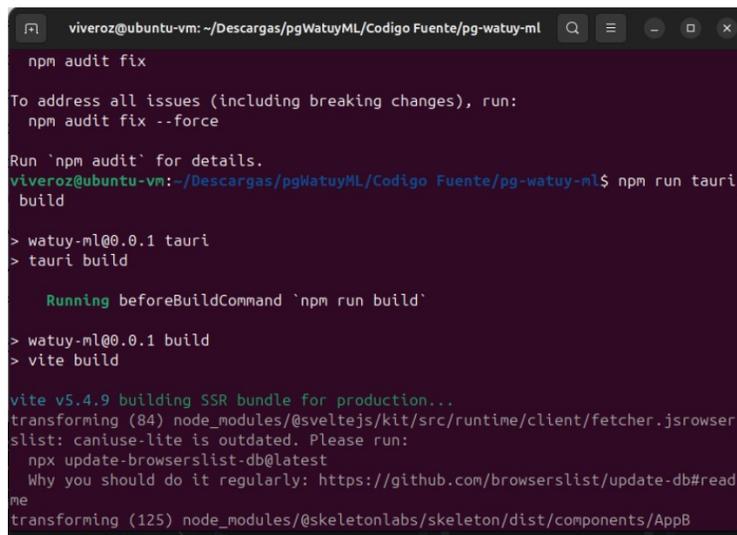


9) Para instalar los paquetes de NodeJS que la aplicación utiliza, use el comando: **npm i**



```
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml$ npm i
( [REDACTED] ): reify: lucide-svelte: timing metavuln:cache:get:security-
```

10) Para compilar la aplicación ejecute el siguiente comando, esto podría tomar un tiempo: **npm run tauri build**



```
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml
npm audit fix
To address all issues (including breaking changes), run:
npm audit fix --force
Run 'npm audit' for details.
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml$ npm run tauri
build
> watuy-ml@0.0.1 tauri
> tauri build

Running beforeBuildCommand `npm run build`
> watuy-ml@0.0.1 build
> vite build

vite v5.4.9 building SSR bundle for production...
transforming (84) node_modules/@sveltejs/kit/src/runtime/client/fetcher.js
browserlist: caniuse-lite is outdated. Please run:
  npx update-browserslist-db@latest
  Why you should do it regularly: https://github.com/browserslist/update-db#readme
transforming (125) node_modules/@skeletonlabs/skeleton/dist/components/AppB
```

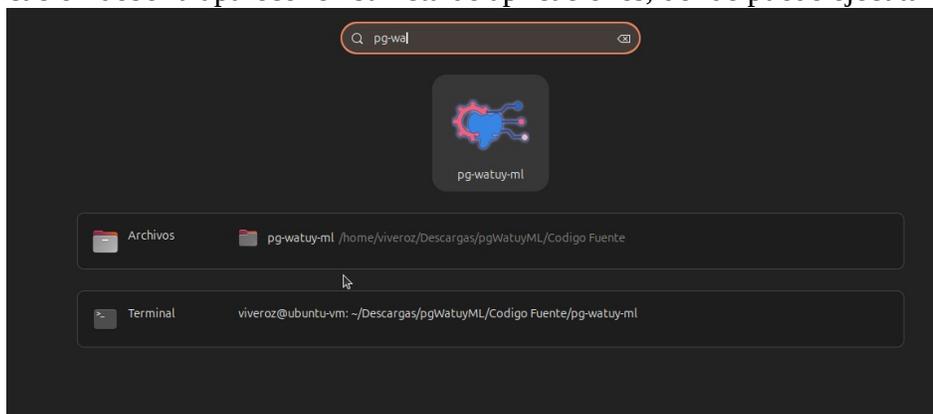
- 11) Tome nota de la ruta en la que se creó el archivo **.deb**, ya que lo necesitará para instalar la aplicación, tenga en cuenta únicamente la parte de la ruta después de **src-tauri**

```
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml
Compiling time-core v0.1.2
Compiling time v0.3.36
Compiling sqlx-macros v0.8.2
Compiling byte-unit v5.1.4
Compiling app v0.1.0 (/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri)
Compiling fern v0.6.2
Compiling csv-core v0.1.11
Compiling csv v1.3.0
Compiling sqlx v0.8.2
Compiling encoding_rs v0.8.35
Finished `release` profile [optimized] target(s) in 2m 28s
Built application at: /home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/app
Bundling pg-watuy-ml_0.1.0_amd64.deb (/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/deb/pg-watuy-ml_0.1.0_amd64.deb)
Bundling pg-watuy-ml-0.1.0-1.x86_64.rpm (/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/rpm/pg-watuy-ml-0.1.0-1.x86_64.rpm)
Bundling pg-watuy-ml_0.1.0_amd64.AppImage (/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/appimage/pg-watuy-ml_0.1.0_amd64.AppImage)
```

- 12) Con el comando anterior se habrá compilado la aplicación, para instalar la aplicación puede utilizar el comando **sudo apt install <ruta a deb>** por ejemplo basado en el paso anterior sería: **sudo apt install ./src-tauri/target/release/bundle/deb/pg-watuy-ml_0.1.0_amd64.deb**

```
viveroz@ubuntu-vm: ~/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml
Compiling sqlx v0.8.2
Compiling encoding_rs v0.8.35
Finished `release` profile [optimized] target(s) in 2m 28s
Built application at: /home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/app
Bundling pg-watuy-ml_0.1.0_amd64.deb (/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/deb/pg-watuy-ml_0.1.0_amd64.deb)
Bundling pg-watuy-ml-0.1.0-1.x86_64.rpm (/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/rpm/pg-watuy-ml-0.1.0-1.x86_64.rpm)
Bundling pg-watuy-ml_0.1.0_amd64.AppImage (/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/appimage/pg-watuy-ml_0.1.0_amd64.AppImage)
Finished 3 bundles at:
/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/deb/pg-watuy-ml_0.1.0_amd64.deb
/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/rpm/pg-watuy-ml-0.1.0-1.x86_64.rpm
/home/viveroz/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml/src-tauri/target/release/bundle/appimage/pg-watuy-ml_0.1.0_amd64.AppImage
viveroz@ubuntu-vm:~/Descargas/pgWatuyML/Codigo Fuente/pg-watuy-ml$ sudo apt install ./src-tauri/target/release/bundle/deb/pg-watuy-ml_0.1.0_amd64.deb
```

13) La aplicación debería aparecer en su lista de aplicaciones, donde puede ejecutarla



14) Puede cerrar la terminal si lo desea.

3. SECCIONES DE LA APLICACIÓN

3.1. Conexión

La aplicación inicia en la pantalla de conexión, donde se puede añadir, modificar y eliminar diferentes conexiones con bases de datos, es fundamental conectarse para acceder a las demás secciones de la aplicación.



- 1) Menú de navegación: En este menú se puede seleccionar la sección de la aplicación a la cual se quiere acceder. En el pantallazo está seleccionada la sección actual (Conexión), mas adelante se describen las demás secciones.
- 2) Añadir conexión: Se puede usar este botón para añadir una nueva conexión, esta se va a seleccionar después de añadirla. A continuación un ejemplo de nueva conexión:

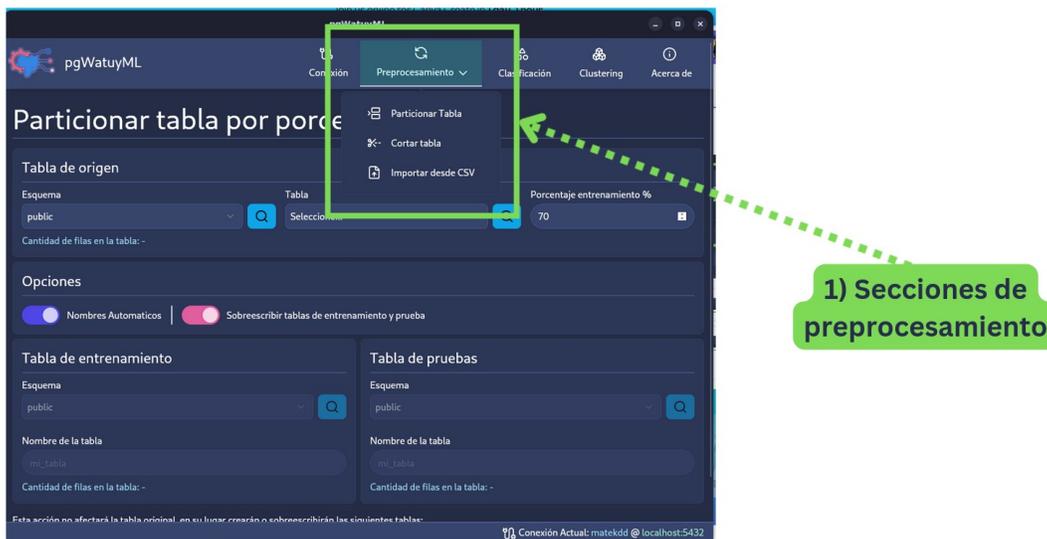
The image shows a close-up of the 'Detalles de la conexión' (Connection Details) form. It has a dark background and contains the following fields and controls:

- Nombre de la conexión: Mi conexión2
- Host: localhost
- puerto: 5432
- Base de datos: matekdd
- Usuario: postgres
- Contraseña: masked with asterisks
- Buttons: Eliminar (trash icon) and Guardar y Conectar (refresh icon)

- 3) Seleccionar conexión: Se puede pulsar sobre el nombre de una conexión para seleccionarla.
- 4) Formulario de conexión: En este formulario se pueden editar la información de la conexión seleccionada. Se tienen los siguientes campos:

- Nombre de la conexión: puede ser cualquier nombre deseado.
 - Host: El host donde se encuentra la base de datos.
 - Puerto: El puerto donde se encuentra la base de datos.
 - Base de datos: El nombre de la base de datos.
 - Usuario: El usuario para conectarse a la base de datos.
 - Contraseña: La contraseña del usuario.
- 5) Botones de acción:
- Botón de Eliminar: Se utiliza para borrar la conexión seleccionada.
 - Botón de Guardar y Conectar: Se utiliza para intentar conectarse a la base de datos, en caso de que la conexión sea exitosa se guardarán los detalles de la conexión.
- 6) Detalles de la conexión: En caso de que actualmente haya una conexión activa, se mostrará los detalles en esta sección (base de datos, host y puerto).

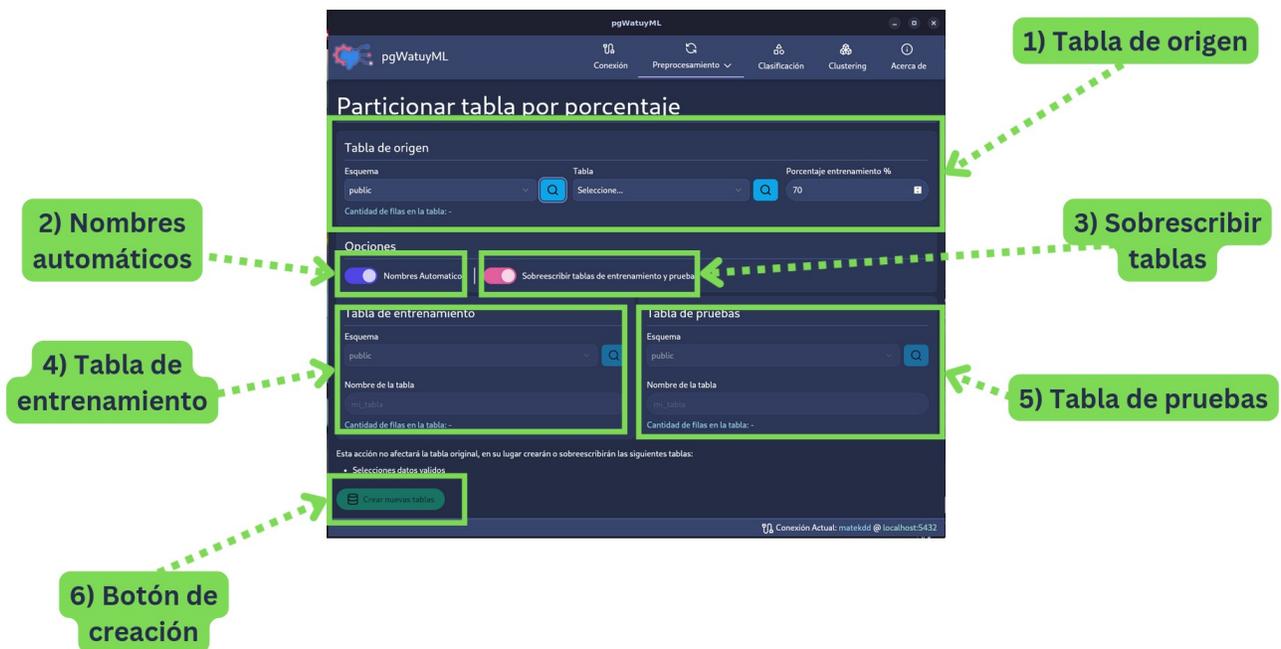
3.2. Preprocesamiento



- 1) Secciones de preprocesamiento: Al pulsar sobre Preprocesamiento, se puede acceder a las diferentes secciones relacionadas, las cuales se describen a continuación.

3.2.1. Particionar Tabla

En esta pantalla se puede seleccionar una tabla de origen, y mediante un porcentaje especificado crear 2 tablas nuevas con los datos de la tabla original en orden aleatorio, estas tablas pueden tener un nombre automático o manual.



- 1) Tabla de origen: La tabla desde la cual se va a copiar las filas, aquí se puede especificar también el porcentaje, este será usado para la tabla de entrenamiento y lo que sobre se irá a la tabla de pruebas.
- 2) Nombres automáticos: Activar esta opción hará que los nombres se creen de manera automática en base a la tabla de origen (Solo para previsualización, no se pueden modificar).

- Nombres Automáticos:



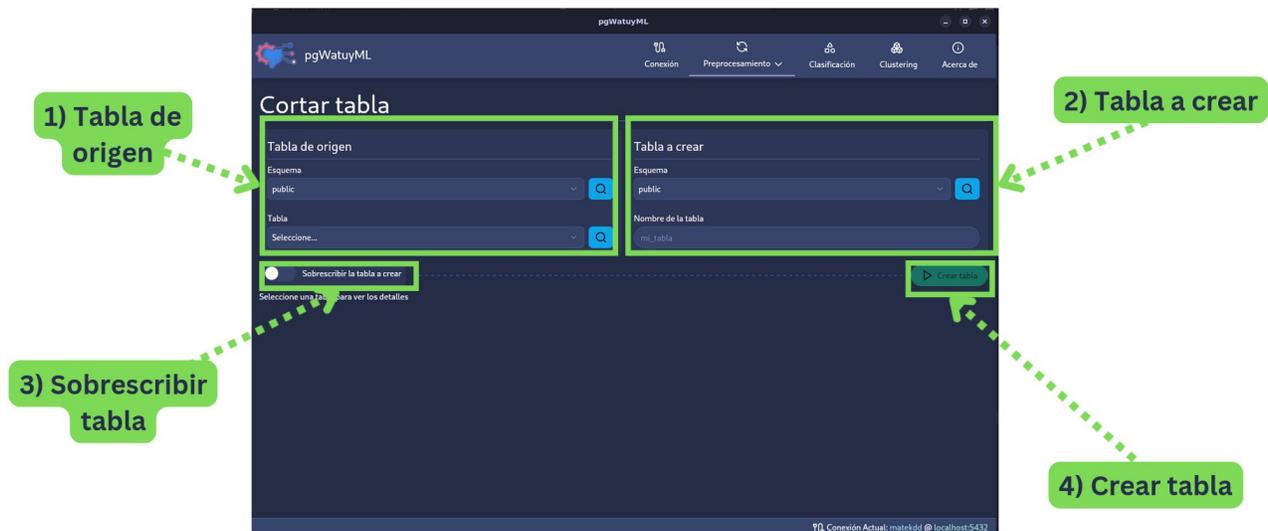
- Nombres Manuales:



- 3) Sobrescribir tablas: Si se activa esta opción y ya existen tablas de pruebas y entrenamiento, estas se sobrescribirán con la nueva información. Si se desactiva esta opción mostrará un error en caso de que ya existan.
- 4) Tabla de entrenamiento: Información de la tabla de entrenamiento (Esquema y nombre).
- 5) Tabla de pruebas: Información de la tabla de pruebas (Esquema y nombre).
- 6) Botón de creación: Este botón se habilita cuando los datos son validos, y se usa para iniciar la operación de la división de la tabla.

3.2.2. Cortar tabla

En esta pantalla se puede elegir una tabla, y luego crear una nueva tabla a partir de las columnas seleccionadas, antes de seleccionar una tabla las siguientes opciones son visibles.



- 1) Tabla de origen: La tabla base para crear la nueva tabla.
- 2) Tabla a crear: Información de la tabla que se va a crear con las columnas seleccionadas.
- 3) Sobrescribir tabla: Si se habilita esta opción al momento de crear la tabla, si esta ya existe se sobrescribirá.
- 4) Crear tabla: Este botón se usa para crear la nueva tabla.

Una vez se seleccione una tabla de origen, se habilitan las siguientes opciones.

1) Columnas

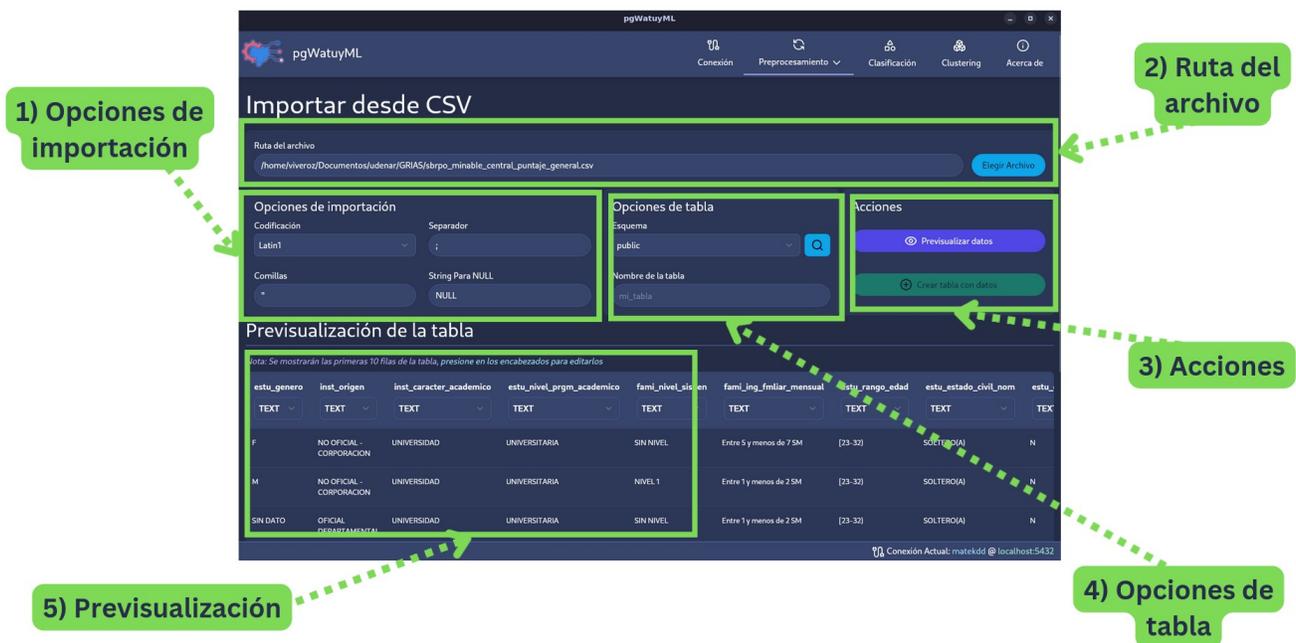
2) Distribución de datos

3) Visualización de datos

- 1) Columnas: En esta sección se pueden ver las columnas de la tabla seleccionada, se puede seleccionar si mantener o no la columna en la nueva tabla, también se puede hacer clic en su nombre para seleccionarla y ver mas información.
- 2) Distribución de datos: En esta tabla se muestran los datos de la columna y su cantidad.
- 3) Visualización de datos: En este grafico se muestra la distribución de datos de la tabla.

3.2.3. Importar desde CSV

En esta pantalla se puede elegir un archivo CSV para crear una tabla e importar los datos ahí.



1) Opciones de importación: Aquí se puede definir opciones como:

- 1) Codificación: La codificación del archivo
- 2) Separados: El separador de columnas
- 3) Comillas: El carácter a utilizar para comillas
- 4) String para NULL: Cadena de caracteres usada para representar datos NULL.

2) Ruta del archivo: Un formulario donde ingresar la ruta del archivo CSV, también cuenta con un botón para elegir el archivo a través del explorador de archivos del sistema.

3) Acciones: Contiene un botón para previsualización los datos y un botón para crear la tabla con los datos. El botón de crear la tabla solo se habilita cuando las opciones sean validas y se haya previsualizado la tabla.

4) Opciones de tabla: En esta sección se define el esquema y nombre de la tabla que se va a crear.

5) Previsualización: En esta parte se muestran las columnas y datos de la tabla, en caso de que se requiera, es posible cambiar el nombre de las columnas y el tipo de columna haciendo click en estos.

3.3. Clasificación

En esta sección se puede crear y probar arboles de decisión usando los algoritmos C4.5 e ID3.

1) Tipo de algoritmo

2) Parámetros del algoritmo

3) Opciones de entrenamiento

4) Opciones de prueba

5) Resultado

The screenshot shows the pgWatuyML interface with the 'Clasificación' (Classification) tab selected. The 'C4.5 ID3' algorithm is chosen. Parameters are set to 'Objetos Mínimos: 1' and 'Factor de Confianza: 0.25'. The training data table is 'public'. The evaluation options are 'Usar la tabla de entrenamiento'. The results section shows a decision tree and evaluation metrics.

```
==== Arbol ====
- jugar_numerico
  - clima = soleado
    - humedad <= 70 -> si - (2) - 100.00%
    - humedad > 70 -> no - (3) - 100.00%
  - clima = nublado -> si - (4) - 100.00%
  - clima = lluvioso -> si - (5 / 2) - 60.00%
```

```
==== Información del Arbol ====
Algoritmo Usado   C45
Tabla             jugar_numerico
Atributo Objetivo jugar
Esquema           public
Número de Nodos  6
Número de Hojas   4
```

```
==== Información de la Evaluación ====
Total de Instancias      14
Instancias Clasificadas Correctamente  12  85.71%
Instancias Clasificadas Incorrectamente  2  14.29%
Estadística Kappa        0.6585
Error Absoluto Medio     0.1714
Raíz del Error Medio Cuadrático  0.2928
Error Absoluto Relativo  37.33%
Raíz del Error Cuadrático Relativo  61.10%
```

```
==== Presión detallada por clase ====
```

6) Botón de empezar

The screenshot shows the pgWatuyML interface with the 'Clasificación' tab selected. The 'C4.5' algorithm is chosen. The training data table is 'jugar_numerico'. The evaluation options are 'Usar la tabla de entrenamiento'. The results section shows a decision tree and evaluation metrics. The 'Empezar' button is highlighted.

```
==== Información del Arbol ====
Algoritmo Usado   C45
Tabla             jugar_numerico
Atributo Objetivo jugar
Esquema           public
Número de Nodos  6
Número de Hojas   4
```

```
==== Información de la Evaluación ====
Total de Instancias      14
Instancias Clasificadas Correctamente  12  85.71%
Instancias Clasificadas Incorrectamente  2  14.29%
Estadística Kappa        0.6585
Error Absoluto Medio     0.1714
Raíz del Error Medio Cuadrático  0.2928
Error Absoluto Relativo  37.33%
Raíz del Error Cuadrático Relativo  61.10%
```

```
==== Presión detallada por clase ====
Clase   TPB   FPR   Precisión   Recall   F-Measure   MCC   ROC-Area   PRC-Area
si      1.000  0.400  0.818      1.000    0.900      0.701  0.933    0.939
no      0.600  0.000  1.000      0.600    0.750      0.701  0.933    0.850
Total Ponderado  0.857  0.257  0.883      0.857    0.846      0.701  0.933    0.907
```

```
==== Matriz de Confusión ====
a b <- Clasificado como
3 2 | a = no
0 9 | b = si
```

- 1) Tipo de algoritmo: Se puede elegir entre usar el algoritmo C4.5 (Mas eficiente y funciona con datos numéricos y nominales), o utilizar ID3 (Solo funciona con datos nominales).
- 2) Parámetros del algoritmo: En esta sección unicamente está disponible para hay un botón para modificar la cantidad de objetos mínimos y el factor de confianza.
- 3) Opciones de entrenamiento: Esta tabla se utilizará para crear el árbol de decisión.

- 4) Opciones de prueba: Incluye varias formas de probar el árbol creado con las opciones de entrenamiento, estas se describen a continuación.
- Usar tabla de entrenamiento: Se utiliza la tabla de entrenamiento para evaluar el árbol.
 - Usar una tabla específica: En esta parte se puede especificar una tabla para evaluar el árbol, también se puede activar el modo de predicción para predecir los valores de una tabla utilizando el árbol, si se activa es posible especificar cuantas instancias mostrar en el resultado (un número muy alto podría hacer que la aplicación se congele).



- Crossfold-Validation: Utiliza el método de crossfold validation para evaluar la tabla de entrenamiento, se puede especificar el número de folds.
 - División por porcentaje: Se divide la tabla de entrenamiento de acuerdo al porcentaje especificado, esas filas se utilizarán para crear el árbol, y el resto para evaluarlo.
- 5) Resultado: Aquí se muestra el resultado (Información de árbol y evaluación).

3.4. Clustering

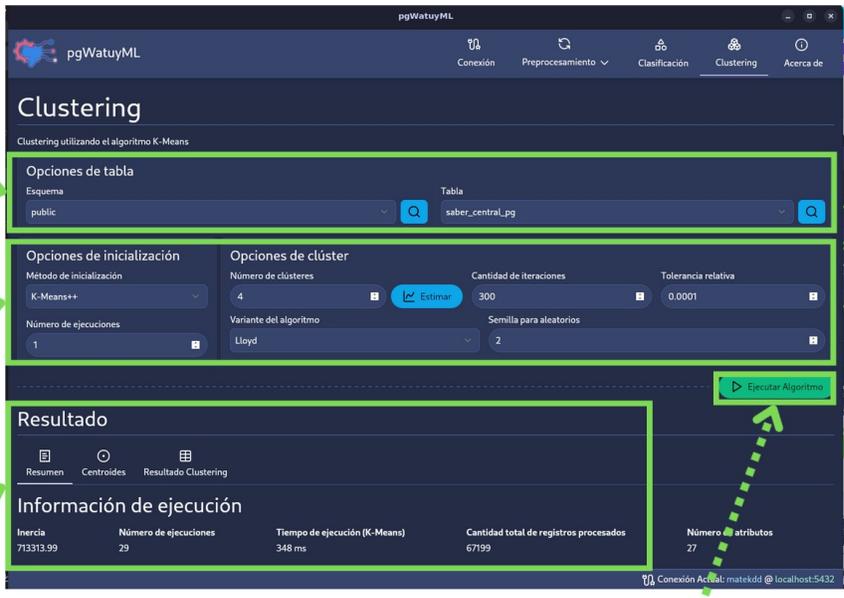
En esta pantalla se pueden entrenar y visualizar la información de clústeres usando el algoritmo K-Means.

1) Opciones de tabla

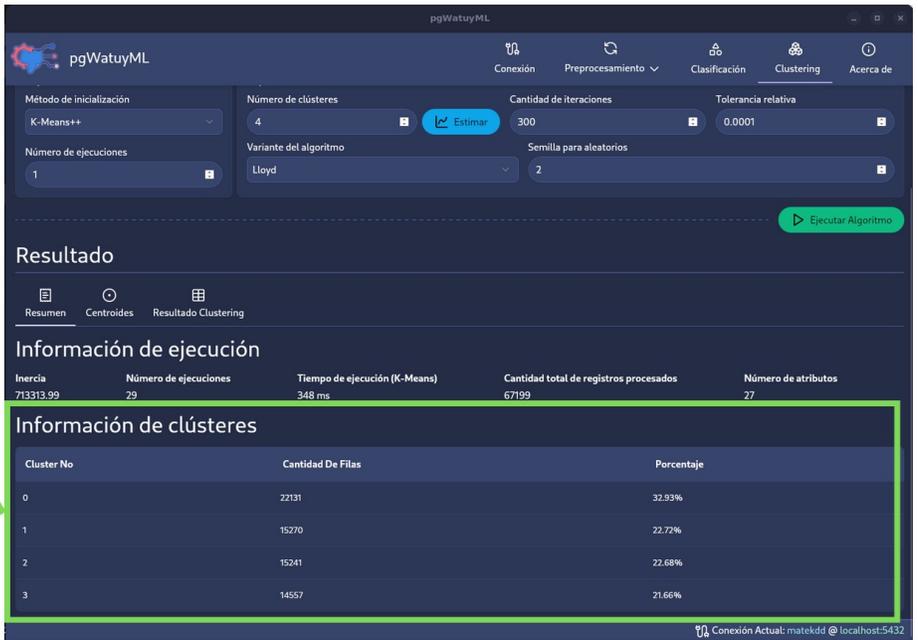
2) Opciones

3) Resultado

4) Ejecutar algoritmo



5) Visualización en tablas



- 1) Opciones de tabla: En esta sección se elige el esquema y la tabla donde se ejecutará el algoritmo.
- 2) Opciones: Se divide en las opciones de inicialización, y opciones de cluster, se puede ajustar para mejorar la efectividad del algoritmo.
 - Opciones de inicialización:

- Metodo de inicialización: Indica el método de inicialización de los centroides
- Número de ejecuciones: Número de veces que se ejecuta el algoritmo k-means con diferentes semillas de centroides. El resultado final es la mejor salida en términos de inercia.
- Opciones de clúster:
 - Número de clusters: Indica el número de clusters a generar. Se puede Estimar utilizando botón Estimar, esto desplegará un menú para utilizar dos metodos:



- Método del codo: Toma como entrada el número máximo de clústeres a evaluar:



Luego muestra el resultado utilizando un grafico de linea.



- Coeficiente de la silueta: Al igual que el método anterior también requiere el número maximo de clústeres a evaluar, pero a diferencia del anterior, se muestra el mejor valor para el número de clústeres y se establece de forma automática.



- Cantidad de iteraciones: Número máximo de iteraciones del algoritmo k-means para una sola ejecución.
 - Tolerancia relativa: Tolerancia relativa con respecto a la diferencia en los centros de los clústeres de dos iteraciones consecutivas para declarar convergencia.
 - Variante del algoritmo: Permite selección la variante del algoritmo K-means a ejecutar.
 - Semilla para aleatorios: Se usa un número entero para que la aleatoriedad sea determinista.
- El botón estimar al lado del número de clusters se puede usar para determinar la mejor cantidad de clusters en la tabla elegida
- 3) Resultado: Aquí se muestra el resultado de los clusters, se puede elegir entre ver el resumen con estadísticas, los centroides y el resultado del clustering, es decir las filas.
 - 4) Ejecutar algoritmo: Este botón se usa para iniciar la creación de los clústeres.
 - 5) Visualización en tabla: En cada sección del resultado se muestra una tabla con la información pertinente.