



Modelos de aprendizaje computacional para la predicción de siniestralidad vial en Bogotá D.C.

Anexo 1: Manual de usuario

Sebastián Rodrigo Castellanos Cardozo

Universidad Antonio Nariño
Facultad de Ingeniería de Sistemas
Bogotá D.C., Colombia
2023

Contenido

- 1 Requerimientos del sistema** **2**

- 2 Configuración del entorno de trabajo** **3**
 - 2.1 Google Drive 3
 - 2.2 Google Colab 6

- 3 Ejecución de los modelos de aprendizaje automático** **10**

Lista de Figuras

2-1	Resultado del navegador web al buscar Google Drive	3
2-2	Resultado del navegador web al buscar Google Colab	4
2-3	Botón para iniciar sesión en Google Drive	4
2-4	Página de inicio después de realizar el inicio de sesión en Google Drive . . .	5
2-5	Página de la carpeta compartida después de acceder desde el link del Paquete de modelos de aprendizaje automático	5
2-6	Resultado del navegador web al buscar Google Colab	6
2-7	Página de inicio de Google Colab	6
2-8	Botón para iniciar sesión en Google Colab	7
2-9	Página de inicio después de realizar el inicio de sesión en Google Colab . . .	7
2-10	Ventana para abrir nuestros cuadernos o <i>notebooks</i>	8
2-11	Cuaderno o <i>notebook</i> <i>'train_rfc_grid.ipynb'</i> abierto en Google Colab	9
3-1	Botón para realizar la ejecución de los cuadernos o <i>notebooks</i>	10
3-2	Ventana para solicitar acceso de Google Colab a Google Drive al realizar la ejecución de los cuadernos o <i>notebooks</i>	11
3-3	Ventana para seleccionar la cuenta de Google Drive a la que deseamos acceder	11
3-4	Ventana para permitir los permisos solicitados de Google Colab a Google Drive	12
3-5	Carpeta 'Resultados' del Paquete de modelos de aprendizaje automático con archivo <i>csv</i> obtenido del modelo RPC (<i>Random Forest Classifier</i>)	13
3-6	Parámetro <i>scoring</i> de la función <i>GridSearchCV</i> usado para obtener el <i>Performance</i> de una métrica de desempeño.	13

1 Requerimientos del sistema

- Sistema operativo: Windows 10 (32 o 64 bits) o superior.
- Software que permita manipular archivos con extensión *csv* como lo pueden ser: Excel, Notepad++, Visual Studio Code, Libre Office, Google Sheets, *etc.*
- Contar con una cuenta de Google, para hacer uso de toda la *suite* de Google Drive y Google Colab.
- Navegador web: Google Chrome, Firefox, Edge, Opera, *etc.*

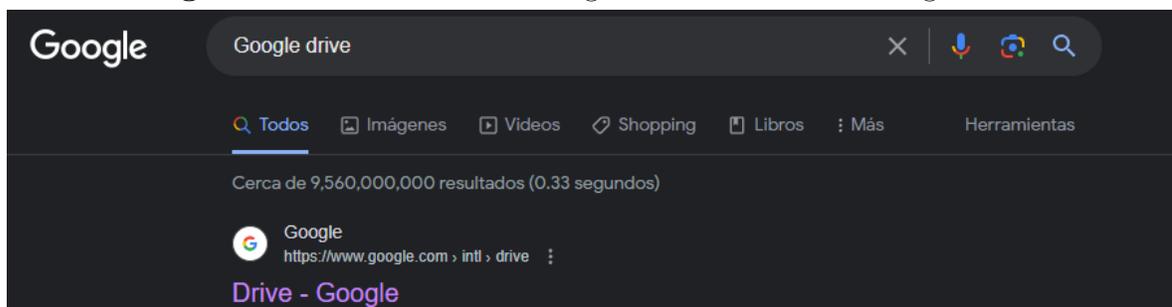
2 Configuración del entorno de trabajo

Antes de comenzar con la configuración del entorno de trabajo, resulta relevante mencionar que se debe tener una cuenta creada en Google para hacer uso de su *suite* de Google Drive y Google Colab. Dicho esto, comenzamos con la configuración de nuestro entorno.

2.1. Google Drive

Para ingresar a Google Drive, debemos entrar a nuestro navegador web de confianza (en mi caso será Google Chrome) y buscamos Google Drive como se puede observar en la Figura 2-1.

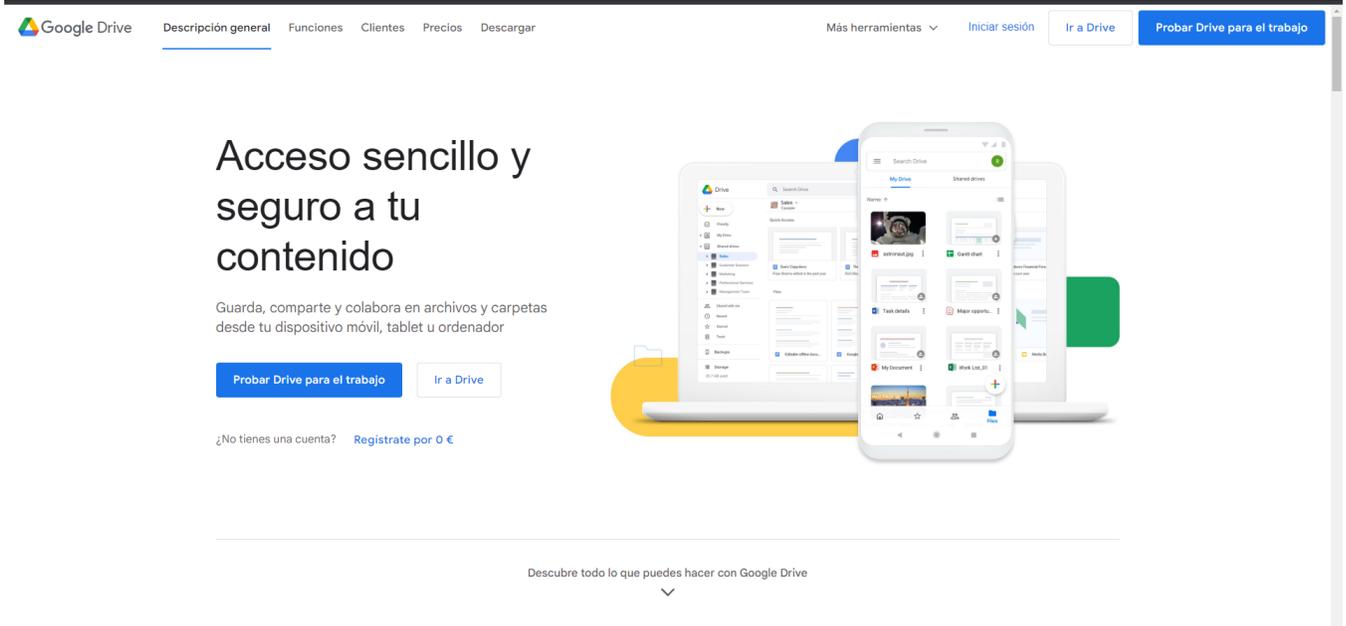
Figura 2-1: Resultado del navegador web al buscar Google Drive



Damos clic en el primer resultado que nos aparece. Al ingresar a Google Drive, se visualizará una pantalla similar a la que se puede observar en la Figura 2-2.

4 CAPÍTULO 2. CONFIGURACIÓN DEL ENTORNO DE TRABAJO

Figura 2-2: Resultado del navegador web al buscar Google Colab

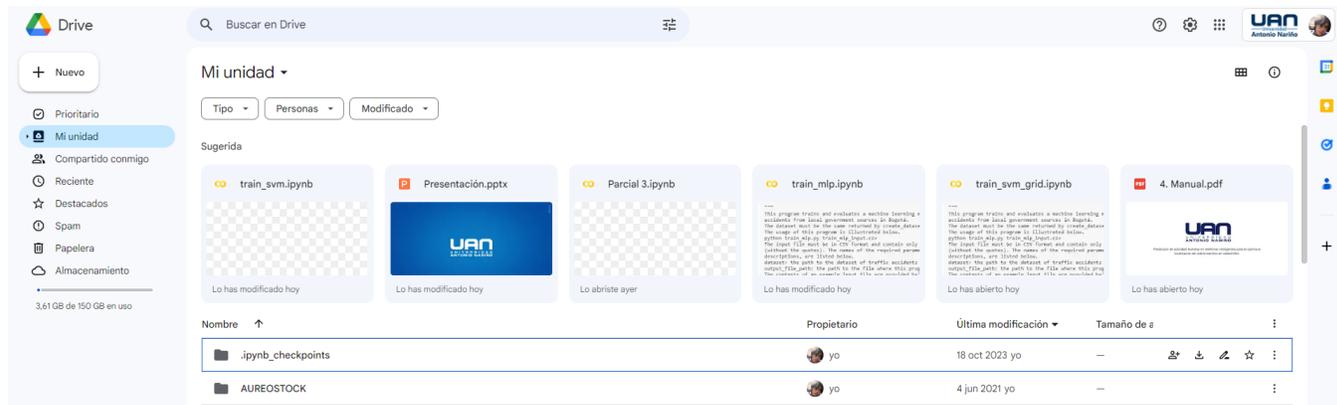


En la parte superior derecha de la Figura 2-2 se puede observar un botón que dice 'Iniciar sesión'. Damos clic en él e iniciamos sesión con nuestra cuenta de Google previamente creada.

Figura 2-3: Botón para iniciar sesión en Google Drive



Al iniciar sesión, la página de Google Drive nos redireccionará a una página similar a la que se observa en la Figura 2-4:

Figura 2-4: Página de inicio después de realizar el inicio de sesión en Google Drive

Después de realizar el inicio de sesión en Google Drive, procedemos a ir a el Paquete de modelos de aprendizaje automático, donde nos redigirá a la siguiente carpeta compartida que se puede observar en la Figura 2-5:

Figura 2-5: Página de la carpeta compartida después de acceder desde el link del Paquete de modelos de aprendizaje automático

Nombre	Propietario	Última modificación	Tamaño de s
CSV	yo	21:46 yo	—
Modelos	yo	21:43 yo	—
Resultados	yo	22:32 yo	—
Scripts	yo	21:45 yo	—

Estando ya en la carpeta del Paquete de modelos de aprendizaje automático, como se observa en la Figura 2-5, podemos observar tres carpetas:

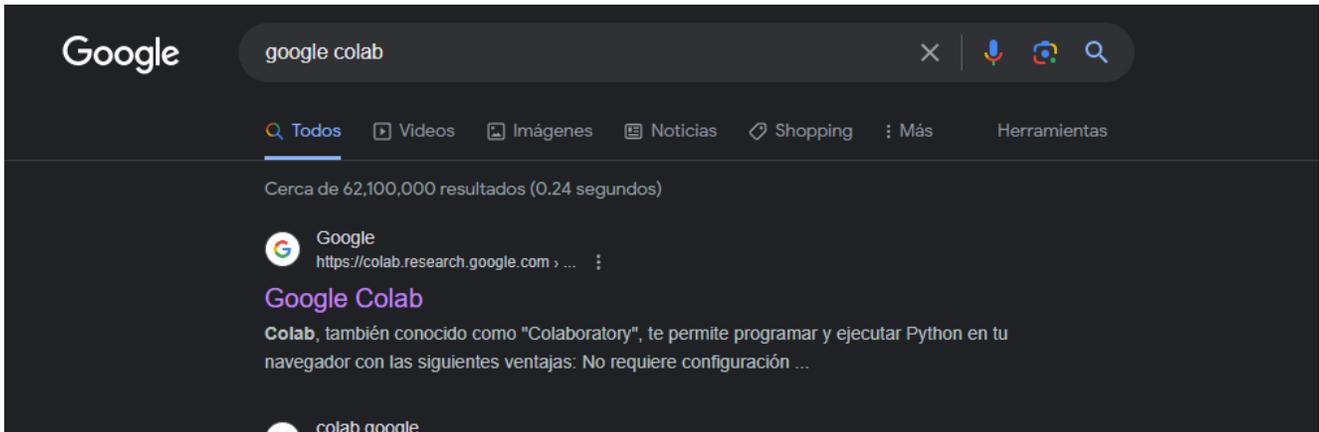
- CSV: Carpeta que contiene los archivos con extensión *csv*, como lo pueden ser los conjuntos de datos, archivos para configurar parámetros, *etc.*
- Modelos: Carpeta que contiene los modelos (cuadernos o *notebooks* de Google Colab en formato *.ipynb*).
- Resultados: Carpeta en la cual se guardaran los resultados de cada modelos de aprendizaje automático.
- Scripts: Carpeta que contiene los *scripts* en formato *.py* necesarios para ejecutar parámetros en los cuadernos o *notebooks* de Google Colab.

Para poder ejecutar los modelos de una manera más simple desde Google Colab, se recomienda clonar la carpeta del Paquete de modelos de aprendizaje automático en la propia cuenta del usuario que desee ejecutar los modelos.

2.2. Google Colab

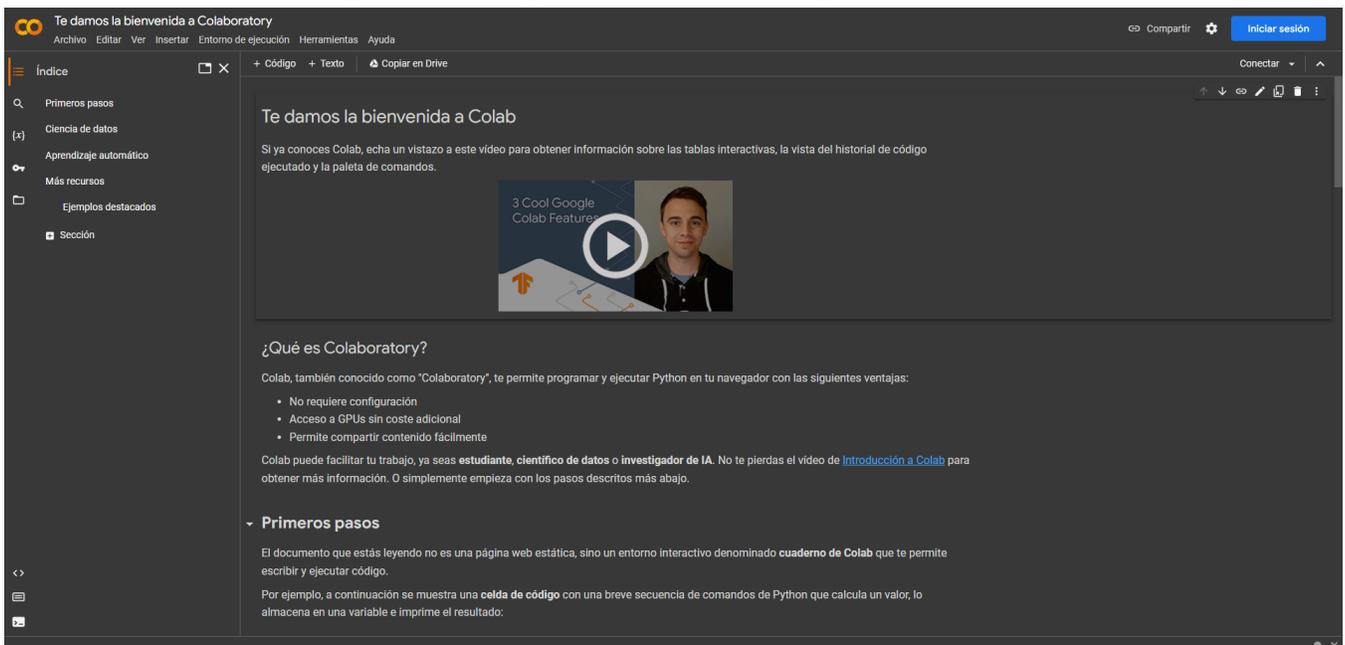
Para ingresar a Google Colab, debemos entrar a nuestro navegador web de confianza (en mi caso será Google Chrome) y buscamos Google Colab como se puede observar en la Figura 2-6.

Figura 2-6: Resultado del navegador web al buscar Google Colab



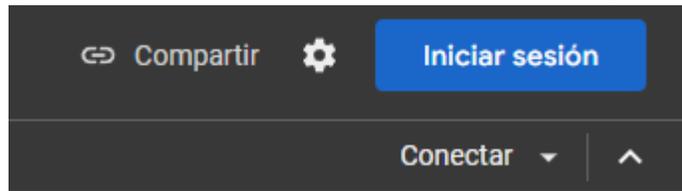
Damos clic en el primer resultado que nos aparece. Al ingresar a Google Colab, se visualizará una pantalla similar a la que se puede observar en la Figura 2-7.

Figura 2-7: Página de inicio de Google Colab



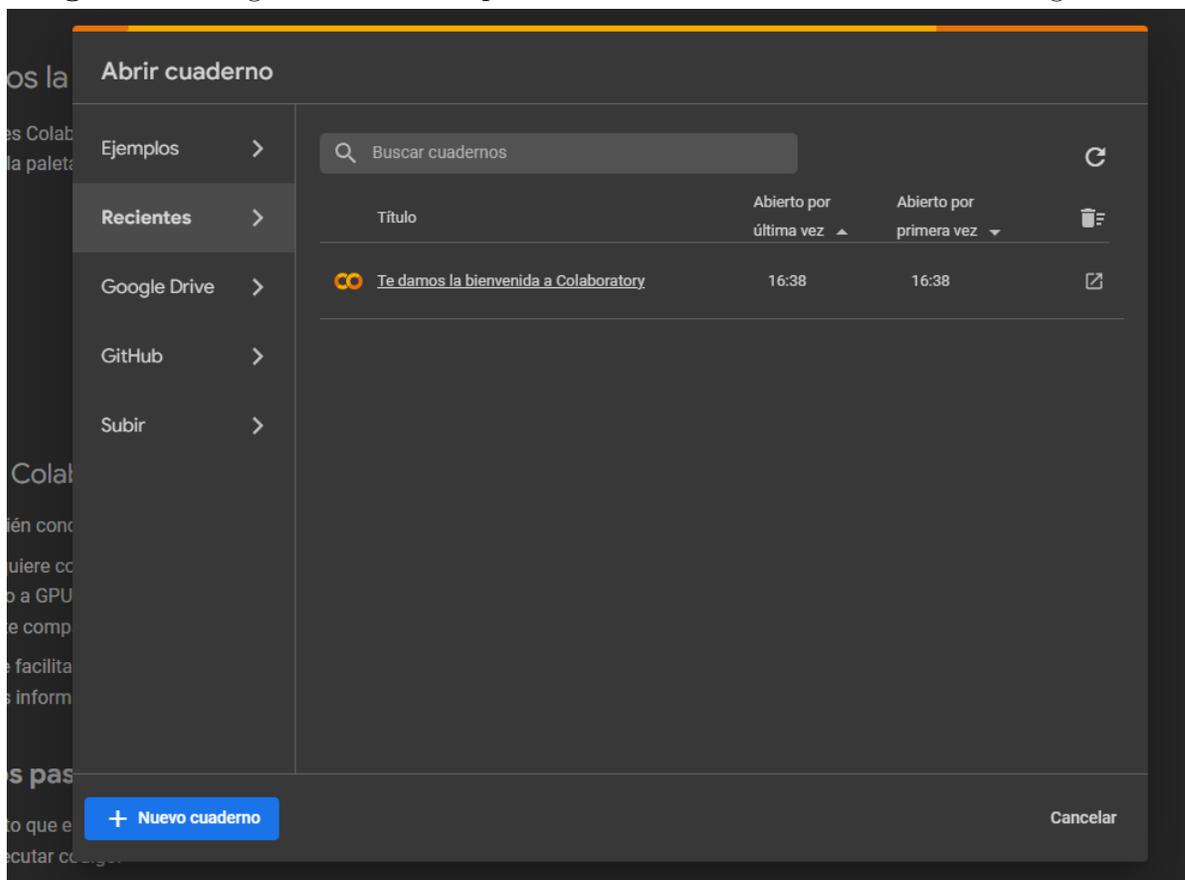
En la parte superior derecha de la Figura 2-7 se puede observar un botón que dice 'Iniciar sesión'. Damos clic en él e iniciamos sesión con nuestra cuenta de Google previamente creada.

Figura 2-8: Botón para iniciar sesión en Google Colab



Al iniciar sesión, la página de Google Colab nos redireccionará a una página similar a la que se observa en la Figura 2-9:

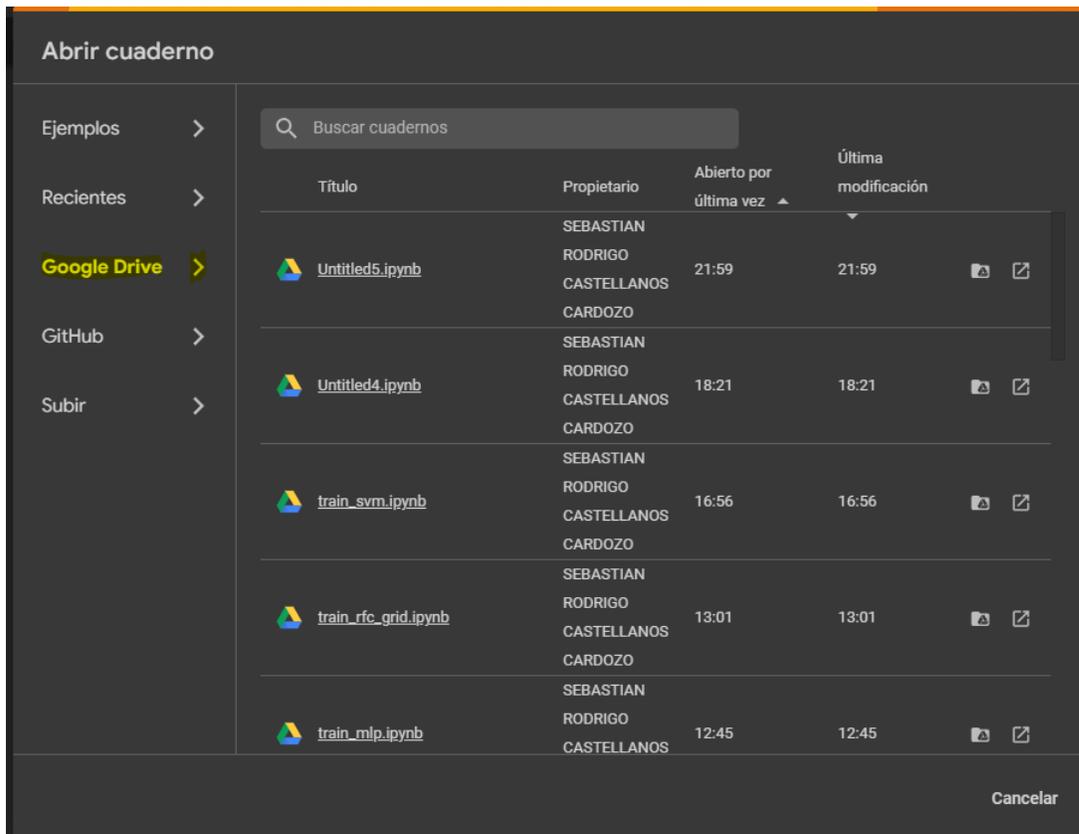
Figura 2-9: Página de inicio después de realizar el inicio de sesión en Google Colab



En la ventana que se puede observar en la Figura 2-9, podemos observar varios elementos. En la parte izquierda, hay un menú donde podemos elegir la fuente para abrir los cuadernos o *notebooks* y *scripts* de Python. En la parte derecha, se pueden visualizar todos los cuadernos o *notebooks* que tengamos creados en esa cuenta. En nuestro caso, al ya realizar la configuración

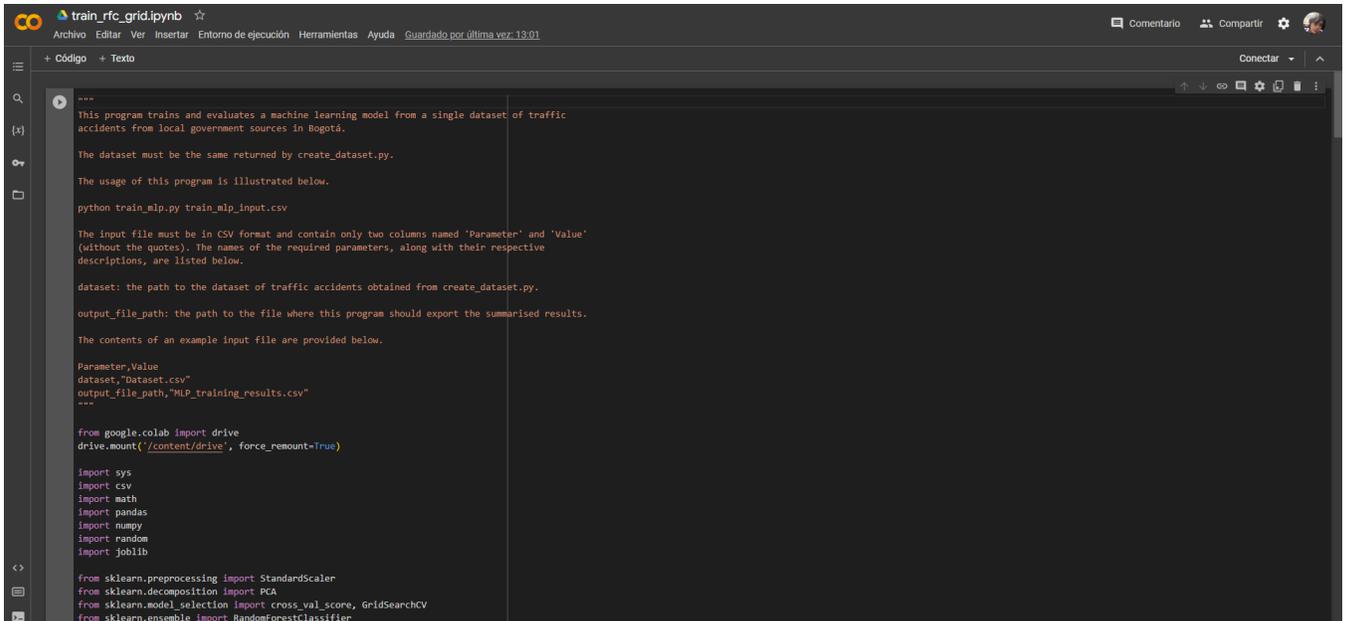
del entorno en Google Drive, seleccionamos en la parte izquierda la opción denominada 'Google Drive', como se observa en la Figura 2-10.

Figura 2-10: Ventana para abrir nuestros cuadernos o *notebooks*



Como se puede observar en la Figura 2-10, al seleccionar la opción 'Google Drive' en la partes derecha se podrán observar todos aquellos cuadernos o *notebooks* que tengamos creados en nuestra cuenta de Google Drive. En este caso, seleccionamos el cuaderno o *notebook* que deseamos abrir. Para este ejemplo, yo seleccionaré el denominado *train_rfc_grid.ipynb* (que es uno de los cuadernos o *notebooks* que se pueden encontrar en la carpeta 'Modelos' del Paquete de modelos de aprendizaje automático).

Al abrir este cuaderno o *notebook*, tendremos algo similar a lo que se puede observar en la Figura 2-11

Figura 2-11: Cuaderno o *notebook* *'train_rfc_grid.ipynb'* abierto en Google Colab

```
train_rfc_grid.ipynb ☆
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Guardado por última vez: 13:01
Comentario Compartir
+ Código + Texto
Conectar
---
This program trains and evaluates a machine learning model from a single dataset of traffic accidents from local government sources in Bogotá.

The dataset must be the same returned by create_dataset.py.

The usage of this program is illustrated below.

python train_mlp.py train_mlp_input.csv

The input file must be in CSV format and contain only two columns named 'Parameter' and 'Value' (without the quotes). The names of the required parameters, along with their respective descriptions, are listed below.

dataset: the path to the dataset of traffic accidents obtained from create_dataset.py.
output_file_path: the path to the file where this program should export the summarised results.

The contents of an example input file are provided below.

Parameter,Value
dataset,"Dataset.csv"
output_file_path,"MLP_training_results.csv"
---

from google.colab import drive
drive.mount('/content/drive', force_remount=True)

import sys
import csv
import math
import pandas
import numpy
import random
import joblib

from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.model_selection import cross_val_score, GridSearchCV
from sklearn.ensemble import RandomForestClassifier
```

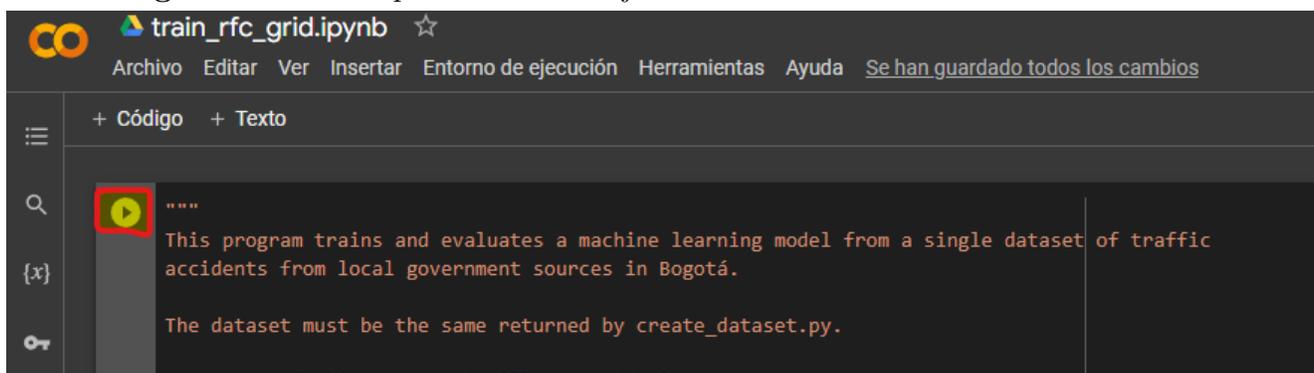
La configuración de nuestros entornos ha terminado. Ahora, como último paso para poder ejecutar de manera correcta los modelos de aprendizaje automático se recomienda realizar los pasos del último capítulo, denominado: Ejecución de los modelos de aprendizaje automático.

3 Ejecución de los modelos de aprendizaje automático

Para la ejecución de los modelos resulta relevante que se hayan completado la configuración del entorno de trabajo de manera correcta. Aclarado esto, se procede a continuar con el proceso de ejecución de los modelos de aprendizaje automático.

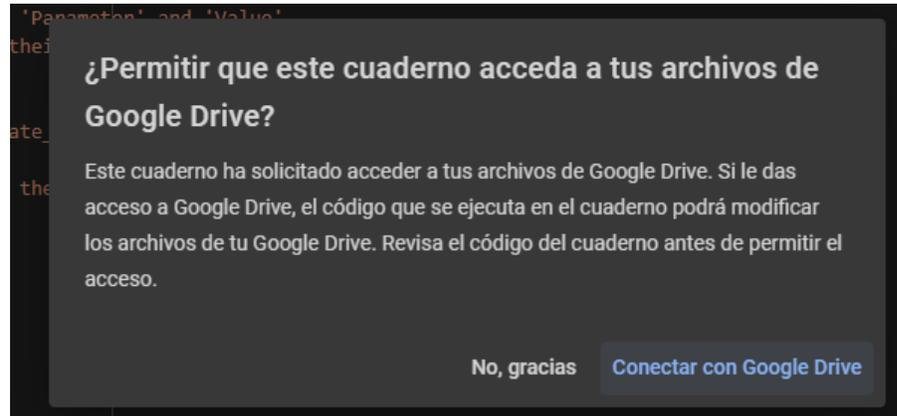
Al tener nuestros cuadernos o *notebooks* abiertos en Google Colab, procederemos a ejecutar los modelos. Para ello, ubicamos el botón que se puede observar en la Figura 3-1

Figura 3-1: Botón para realizar la ejecución de los cuadernos o *notebooks*



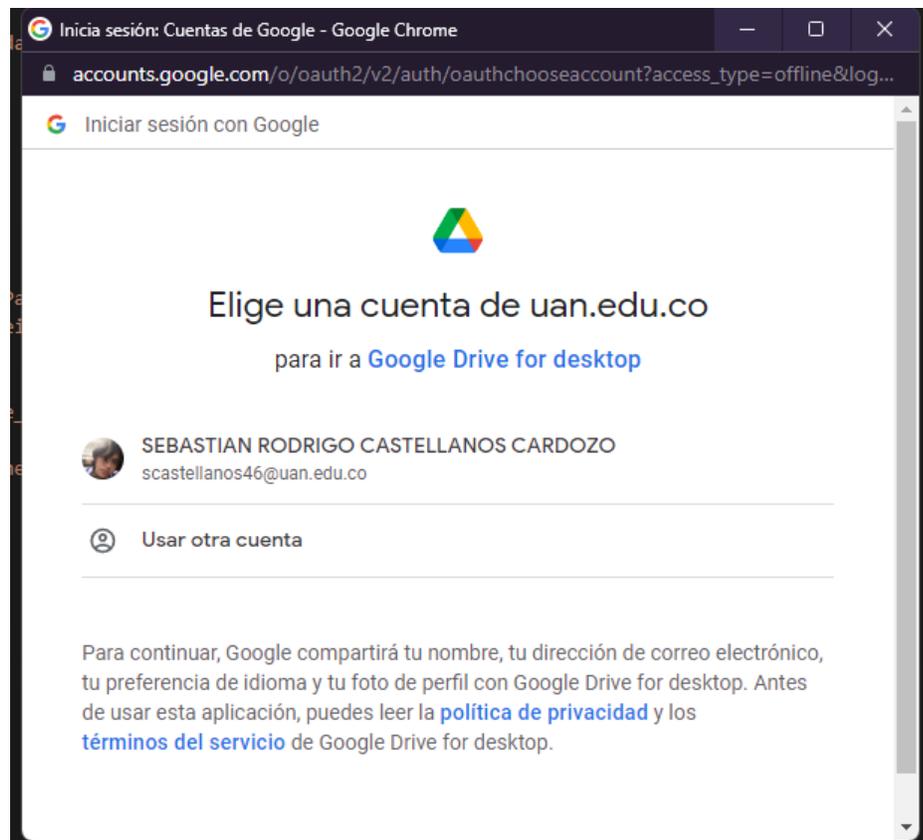
Al darle clic al botón que se puede observar en la Figura 3-1, nos aparecerá la siguiente ventana que se puede observar en la Figura 3-2.

Figura 3-2: Ventana para solicitar acceso de Google Colab a Google Drive al realizar la ejecución de los cuadernos o *notebooks*



Al aparecer la ventana de la Figura 3-2, damos clic en el botón 'Conectar con Google Drive'. Al realizar el clic, se nos abrirá la siguiente ventana que se puede observar en la Figura 3-3.

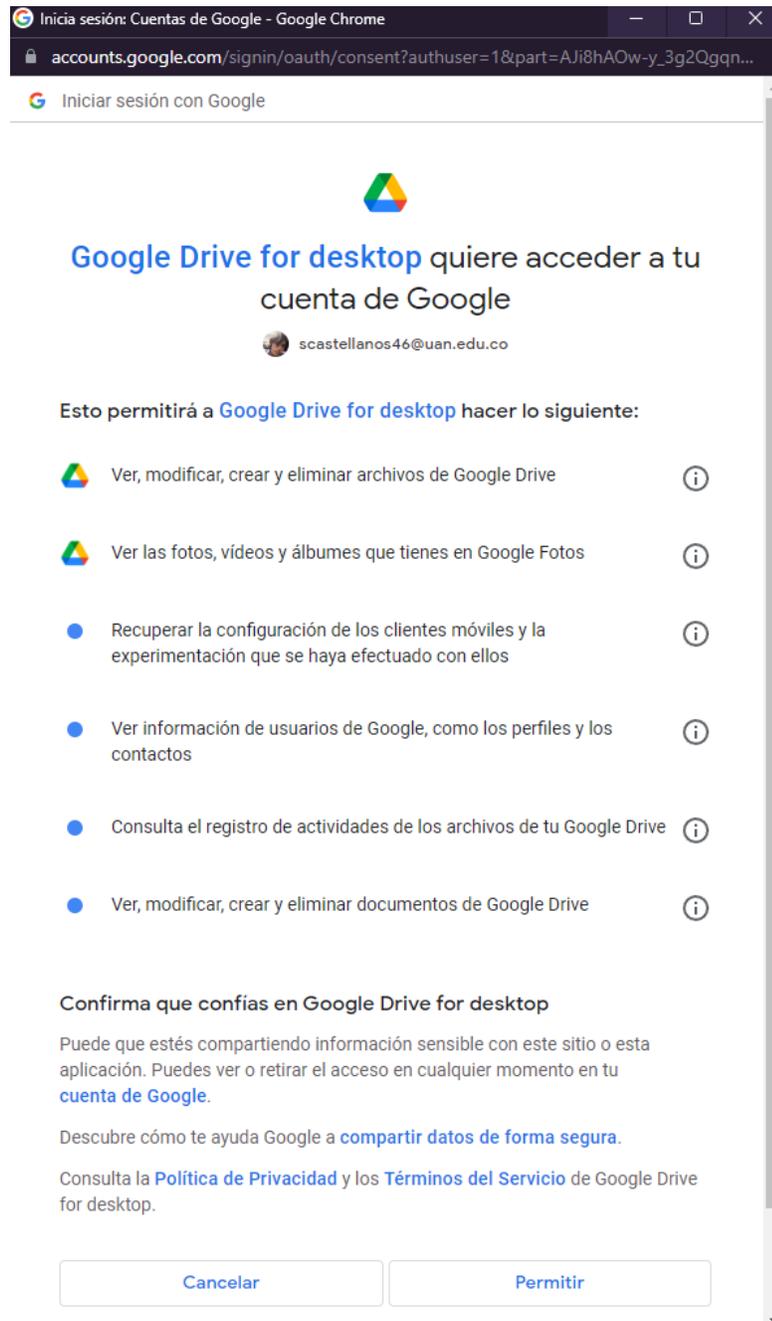
Figura 3-3: Ventana para seleccionar la cuenta de Google Drive a la que deseamos acceder



Damos clic en la cuenta donde realizamos la configuración de los entornos. Al realizar clic,

se nos abrirá la siguiente ventana con los permisos que solicita Google Colab a Google Drive para poder acceder a su contenido. Esto se puede observar en la Figura 3-4.

Figura 3-4: Ventana para permitir los permisos solicitados de Google Colab a Google Drive



Damos clic en el botón 'Permitir' ubicado en la parte inferior derecha de la ventana que se puede observar en la Figura 3-4 para darle acceso total a Google Colab de nuestro contenido almacenado en Google Drive. Al realizar esto, nuestro modelo empezará a ejecutarse. Solo

es cuestión de tiempo para que el modelo finalice y podamos ver los resultados obtenidos en la carpeta 'Resultados' del Paquete de modelos de aprendizaje automático. Tendremos un archivo similar al que se puede observar en la Figura 3-5.

Figura 3-5: Carpeta 'Resultados' del Paquete de modelos de aprendizaje automático con archivo *csv* obtenido del modelo RPC (*Random Forest Classifier*)



Adicional a esto, si se desea obtener el *Performance* de una métrica de desempeño en concreto, puede modificarla a su gusto en el parámetro *scoring* de la función *GridSearchCV*. Para este ejemplo, se usó la métrica de desempeño F1 Score, y en el parámetro *scoring* esta sería 'f1', tal como se puede observar en la Figura 3-6.

Figura 3-6: Parámetro *scoring* de la función *GridSearchCV* usado para obtener el *Performance* de una métrica de desempeño.

```
GridSearchCV(classifier, params, n_jobs = -1, scoring = 'f1', cv = 10)
```

A continuación, podrá observar los parámetros admitidos por el parámetro *scoring* para evaluar al modelo con una métrica de desempeño en concreto.

Nota 1: En el parámetro *scoring* si no se coloca parámetro alguno, por defecto evaluará el modelo con la métrica de desempeño *accuracy*.

- F1 Score: `scoring = 'f1'`.
- Accuracy: `scoring = 'accuracy'`.
- Precision: `scoring = 'precision'`.
- Recall: `scoring = 'recall'`.

Nota 2: Para todos los modelos se debe realizar el mismo proceso de Ejecución de los modelos de aprendizaje automático.