

Herramienta Intelligent Data Analysis Tool

Guía de instalación y ejecución.

La herramienta está construida en lenguaje python, dedicada para el sistema operativo Windows. La instalación de python se puede realizar siguiendo los pasos que se mencionan a continuación.

Windows

Puedes descargar la versión 3.10.7 de la distribución oficial de Python desde el sitio web oficial: <https://www.python.org/downloads/>. Después de descargar el archivo *.exe, debés ejecutarlo y seguir las instrucciones.

NOTA: Es importante instalar **pip** y **tkinter** en este procedimiento.

Instalación herramienta

Descargar la herramienta en el repositorio del siguiente link <https://github.com/Diego-Cansino/Intelligent-Data-Analysis-Tool>, una vez instalada la herramienta en la ruta de su elección, acceder a ésta por medio de la terminal correspondiente a sus sistema operativo y escribir lo siguiente.

```
$> pip install -r requirements.txt
```

Al terminar de realizar la instalación de las librerías contenidas en “requirements.txt” procederemos a la ejecución del programa, para ésto, escriba la siguiente instrucción en la terminal.

```
$> python main.py
```

Manual de usuario.

Intelligent Data Analysis Tool es una herramienta que fue creada con el propósito de gestionar dataset, haciendo limpieza de los datos, así mismo como mostrar gráficas que ayuden al usuario a comprender de manera más clara los datos. Por otra parte, esta aplicación tiene la capacidad de implementar un modelo de aprendizaje automático según los datos ofrecidos por el dataset seleccionado.

La herramienta permitirá al usuario seguir la metodología CRISP-DM, ayudando de esta manera a lograr una buena comprensión de los datos, preparación de los datos y modelado de los datos.

Con lo antes mencionado, podemos entender el alcance que puede llegar a tener esta herramienta, esta puede considerarse una alternativa a la aplicación IBM SPSS Modeler, donde tendremos la capacidad de implementar y/o utilizar aplicaciones que consideremos pertinentes para obtener buenos resultados.

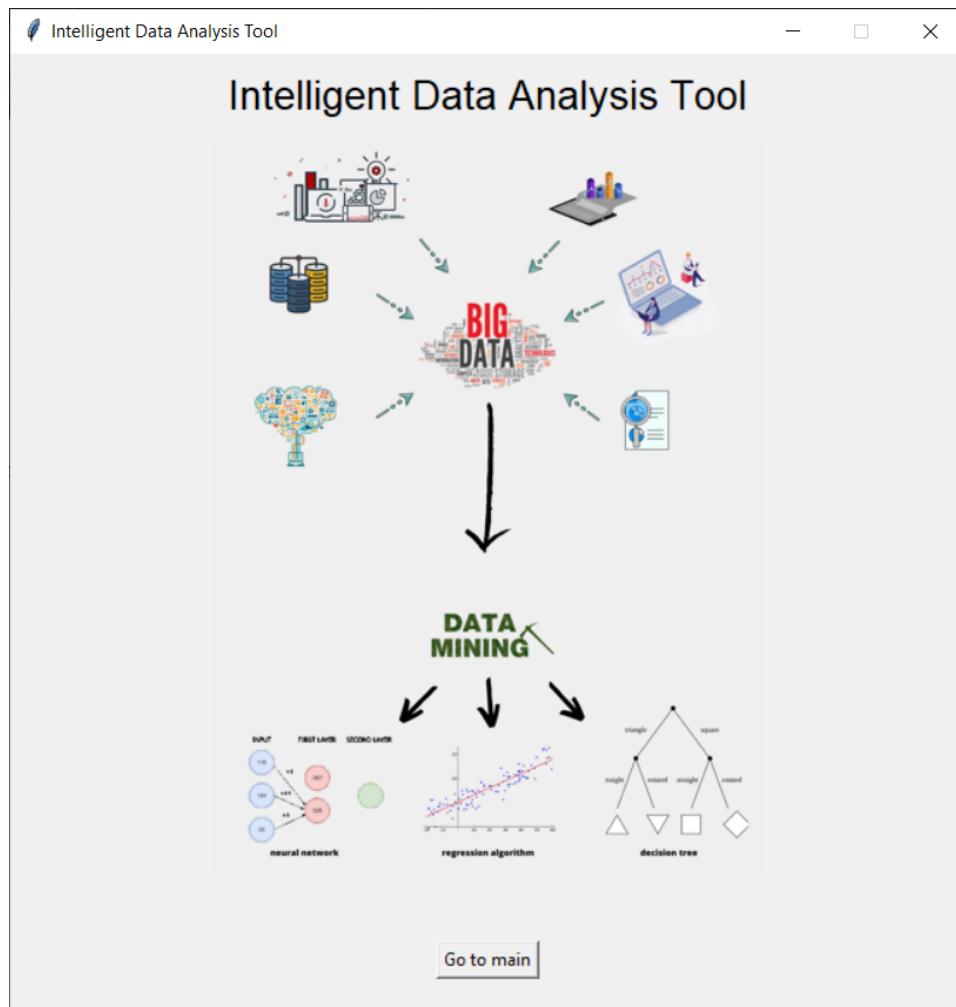


Figura 1.0: Interfaz de inicio

Comprensión de datos

Una vez que seleccionamos el botón “Go to main” accedemos a la interfaz principal, donde tendremos una sección en la cual podremos visualizar nuestros datos, así como una sección llamada “Data understanding”, donde podremos seleccionar el dataset con el que deseemos trabajar, por otra parte podremos graficar los datos de la tabla según nuestras necesidades.

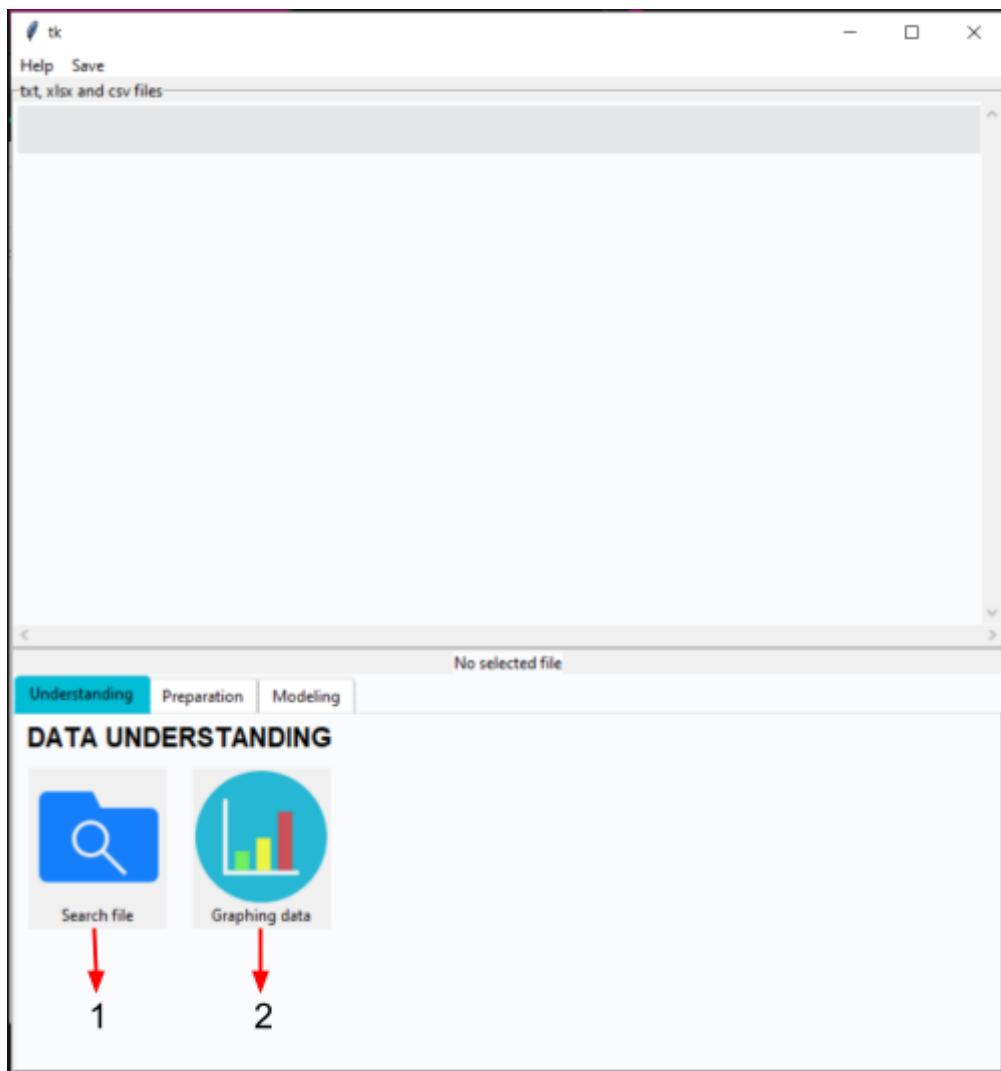


Figura 2.0: Interfaz principal

Interfaz principal

1. “**Search file**” nos permite ingresar un archivo de extensión (txt, xlsx y csv) y lo carga en nuestro visualizador de datos.
2. “**Graphing data**” Parte encargada de graficar los datos cargados.

Selección de dataset

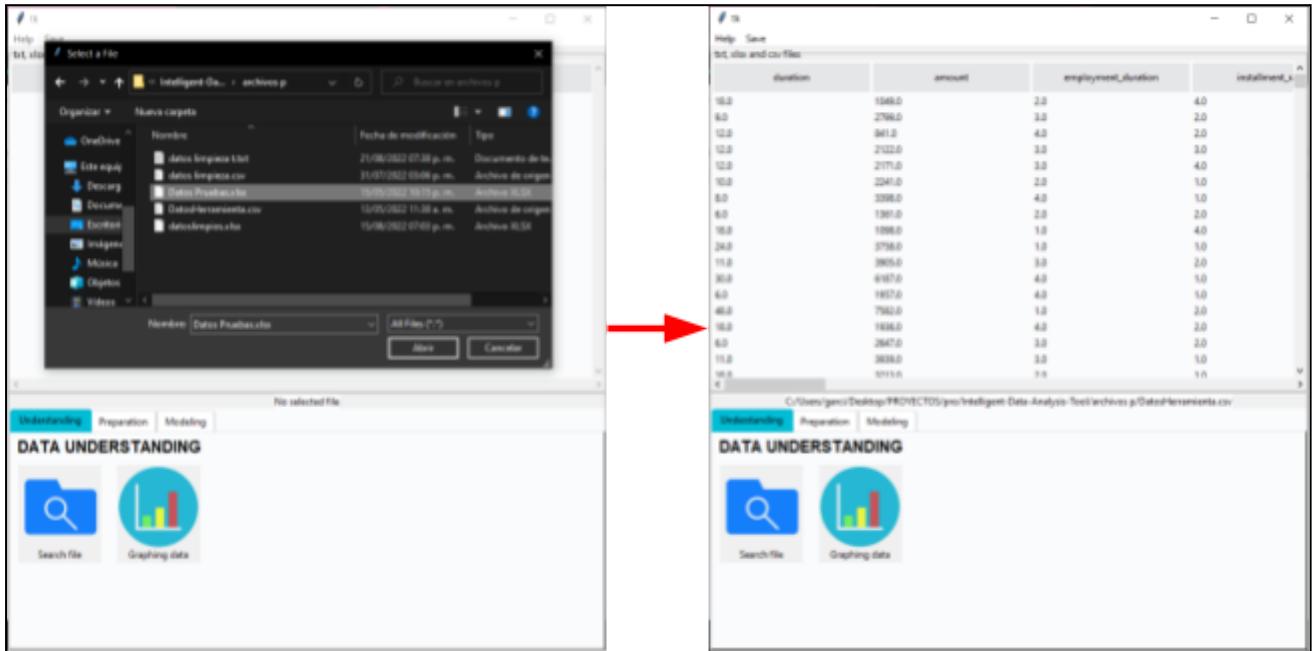


Figura 2.1.1: Selección de dataset

Al encontrar el dataset con el que se desea trabajar, lo seleccionaremos y posteriormente desplegará la información respectiva al dataset que seleccionamos. A continuación en la Figura 2.1.2 podemos comprobar el correcto funcionamiento de la carga de los datos.

DatosHerramienta

Archivo Editar V Help Save
txt, xlsx and csv files

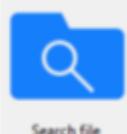
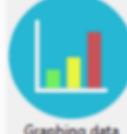
duration amount employment_duration installment_n

A1	duration	amount	employment_duration	installment_n
1	duration	amount		
2	18.000	12.0		
3	9.000	12.0		
4	12.000	12.0		
5	12.000	12.0		
6	12.000	12.0		
7	10.000	12.0		
8	8.000	24.0		
9	6.000	11.0		
10	18.000	30.0		
11	24.000	6.0		
12	11.000	48.0		
13	30.000	18.0		
14	6.000	6.0		
15	48.000	11.0		
16	18.000	18.0		
17	6.000	2		
18	11.000	3		
19	18.000	3		
20	36.000	2		
21	11.000	7		
22	6.000	3		
23	12.000	3		
24	36.000	2		
25	12.000	1		
26	6.000	4		
27	11.000	4		
28	12.000			
29	9.000	1		
30	15.000	3		
31	42.000	4		
32	30.000	3		

C:/Users/garci/Desktop/PROYECTOS/pro/Intelligent-Data-Analysis-Tool/archivos p/DatosHerramienta.csv

Understanding Preparation Modeling

DATA UNDERSTANDING

Search file Graphing data

Figura 2.1.2: Comprobación de datos cargados

Como podemos ver en la figura 2.1.2 podemos rectificar el correcto funcionamiento de la aplicación al cargar datos.

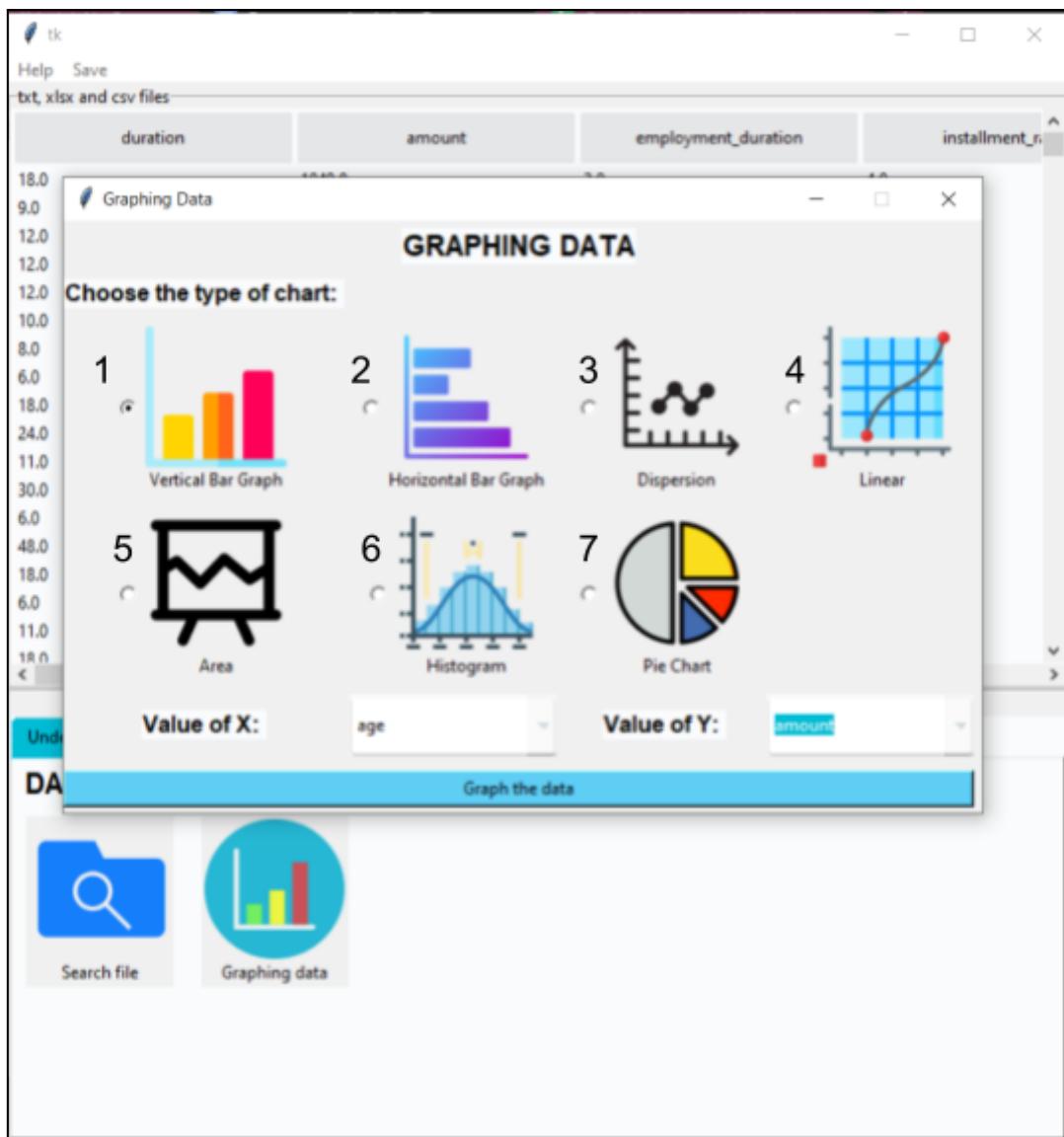


Figura 2.2.1: Selección de tipo de gráfica

Una vez el usuario seleccione “Graphing data” en la interfaz principal se desplegará la interfaz de la “Figura 2.2.1”, es esta sección contamos con lo siguiente:

1. **Gráfica de barras verticales.**
2. **Gráfica de barras horizontales.**
3. **Gráfica de dispersión.**
4. **Gráfica lineal.**
5. **Área.**
6. **Histograma.**
7. **Gráfica de pie.**

Por otra parte, tenemos una sección en la que elegiremos dos campos respectivos a graficar:

- Value of X:** En esta sección agregaremos el campo que deseemos graficar correspondiente al eje de las X
- Value of Y:** En esta sección agregaremos el campo que deseemos graficar correspondiente al eje de las Y

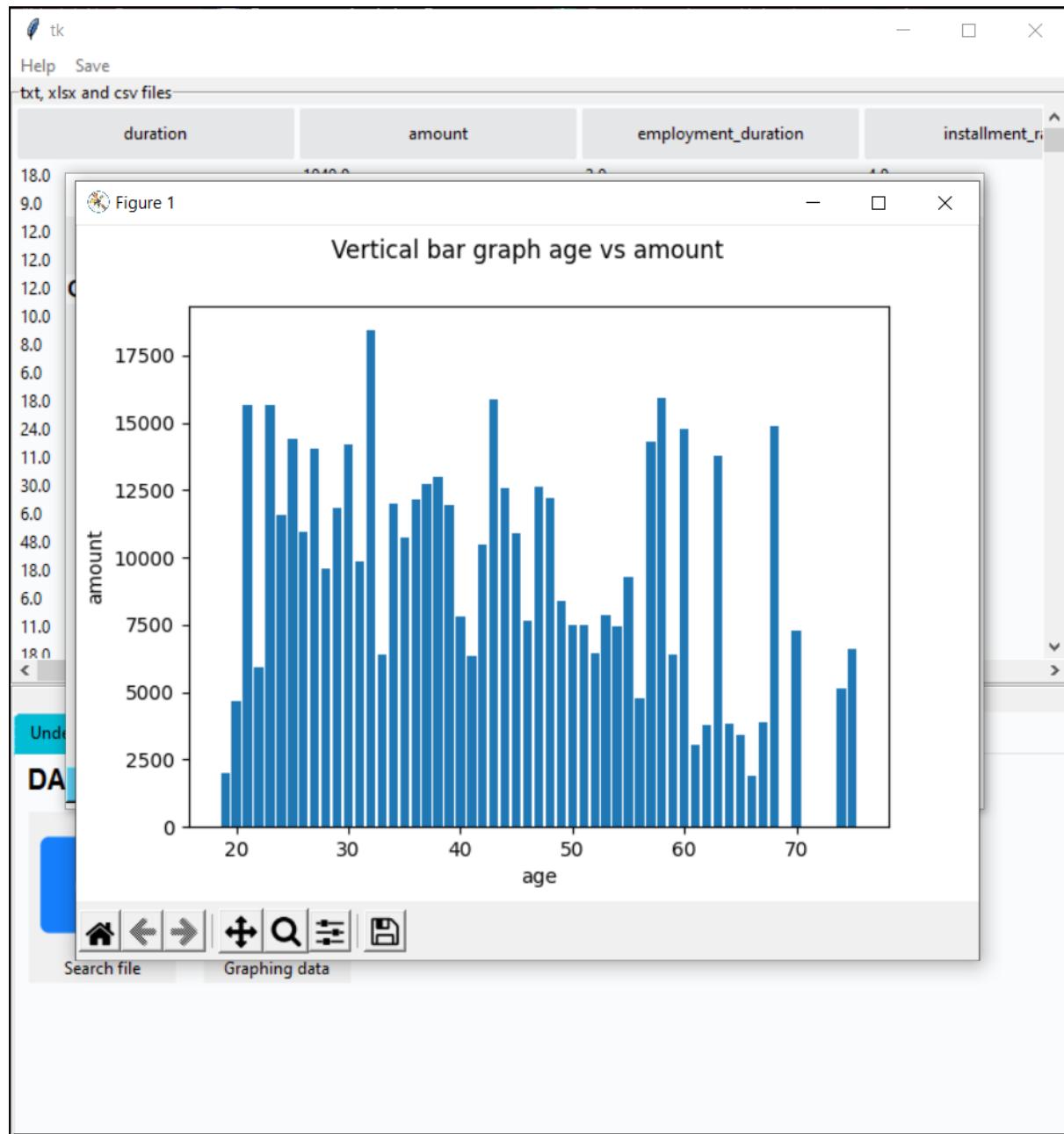


Figura 2.2.2: Despliegue de gráfica

Una vez el usuario detecte los campos que desee utilizar respecto a eje de las Y y eje de las X, así como el tipo de gráfico que desee utilizar, seleccionamos “Graph the data (Figura 2.2.1)”, Al realizar dicho procedimiento se desplegará el gráfico correspondiente a los datos seleccionados, como se muestra en la “Figura 2.2.2”

Preparación de los datos

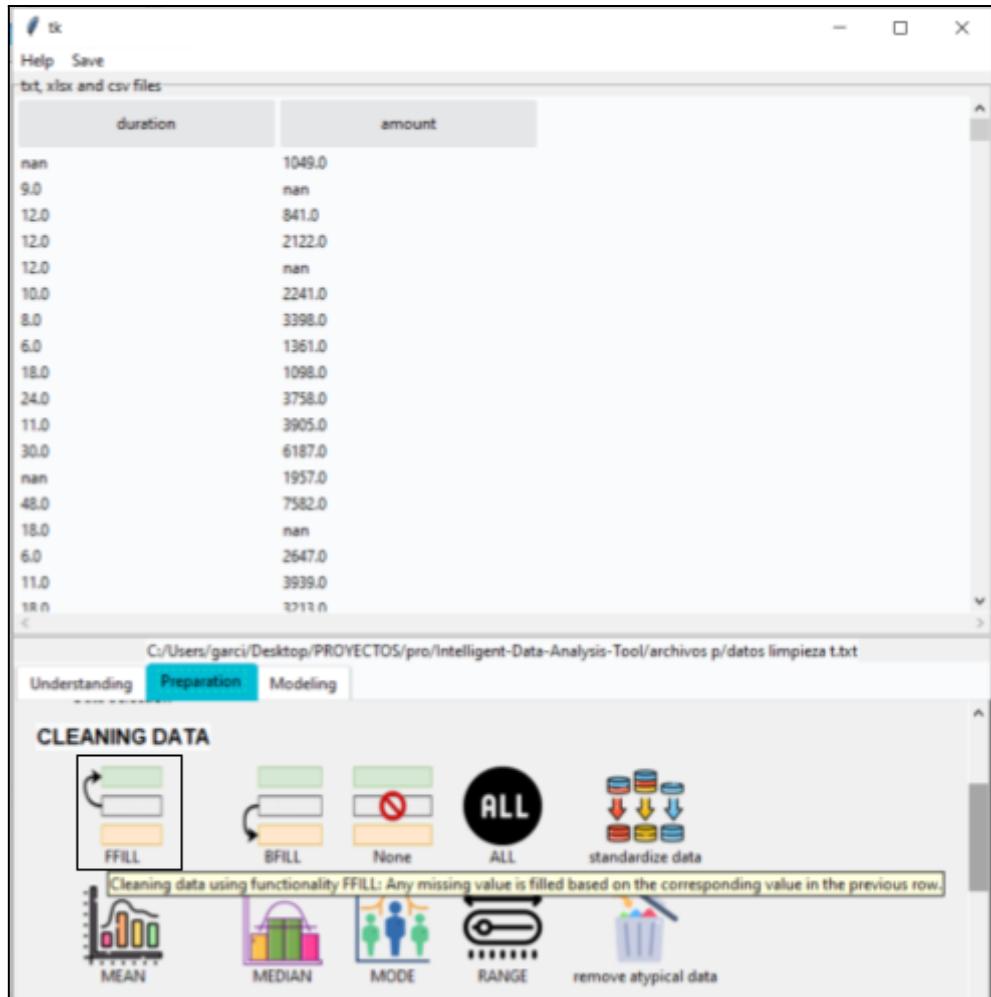


Figura 3.0: Limpieza de datos.

Para la limpieza de los datos se cuenta con 10 tipos diferentes de métodos. Al situarnos sobre alguno de los métodos este nos da una breve retroalimentación de ese método de limpieza, como se muestra en la "Figura 3.0"

Los métodos de limpieza con los que se cuenta son:

1. **FFILL** → Llena los datos vacíos con el valor correspondiente al registro previo
2. **BFILL** → Llena los datos vacíos con el valor correspondiente al siguiente registro
3. **None** → Elimina el registro que cuenta con campos nulos.
4. **ALL** → Combinación de los métodos FFILL, BFILL, None.
5. **Standardize data** → Estandariza datos atípicos.
6. **MEAN** → Llena campos nulos haciendo uso del método estadístico de promedio.

7. **MEDIAN** → Llena campos nulos haciendo uso del método estadístico de mediana.
8. **MODE** → Llena campos nulos haciendo uso del método estadístico de moda.
9. **RANGE** → Llena campos nulos haciendo uso del método estadístico de rango.
10. **Remove atypical data** → Remueve los datos atípicos.

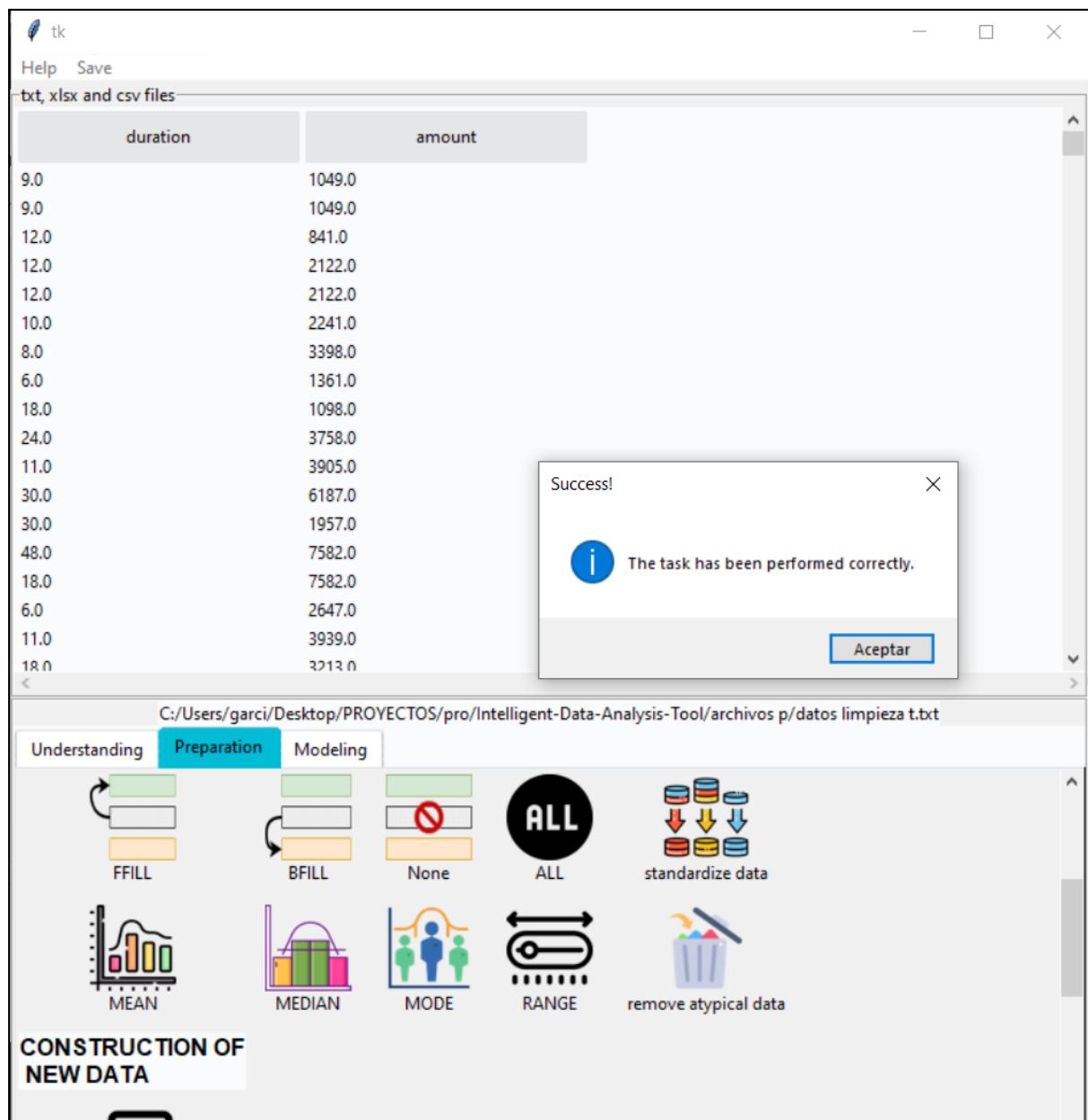


Figura 3.1 Método de limpieza realizado.

Al aplicar un método de limpieza se muestra que la tarea fue lograda con éxito, como en la “Figura 3.1”

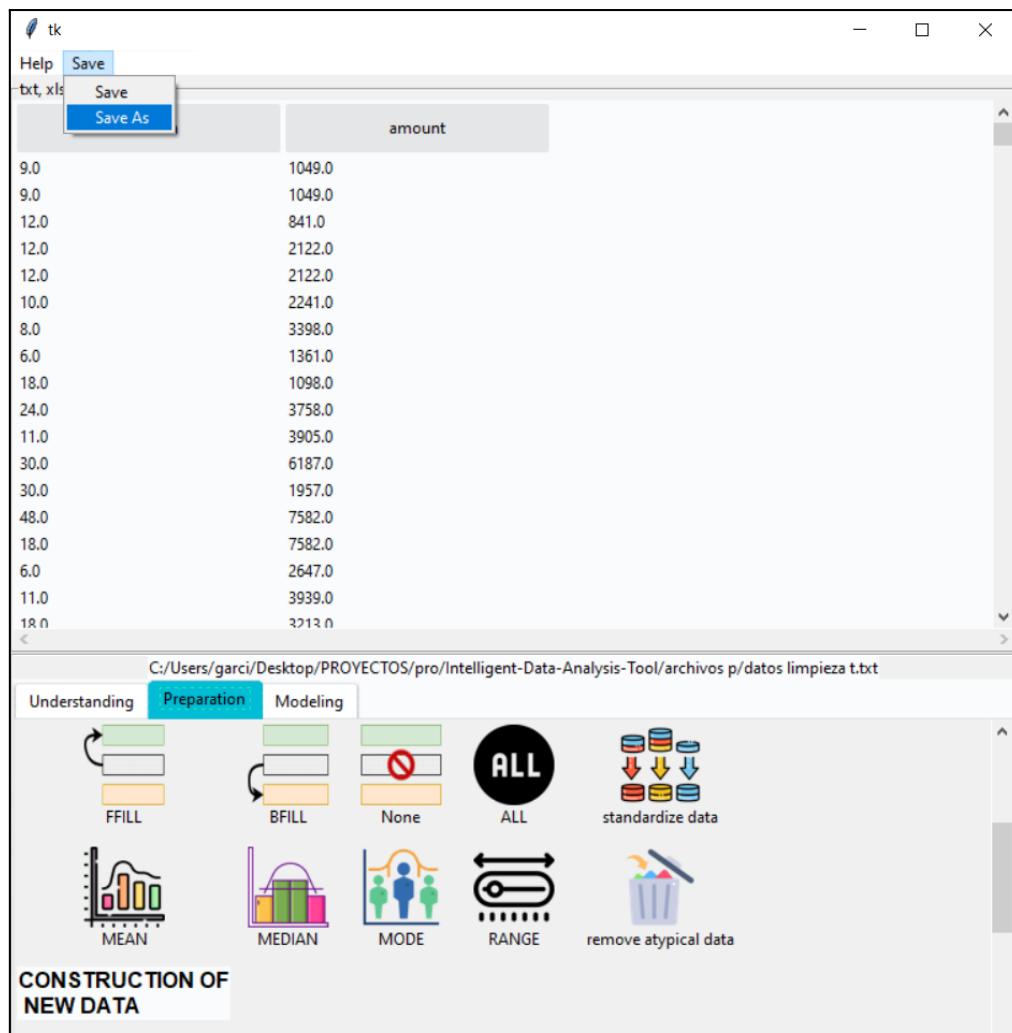


Figura 4.0: Guardar archivos.

Una vez se hayan efectuado tantas modificaciones fueran necesarias sobre el dataset seleccionado, se puede guardar el proyecto. Para ello tenemos la opción de Save. Como se muestra en la "Figura 4.0"

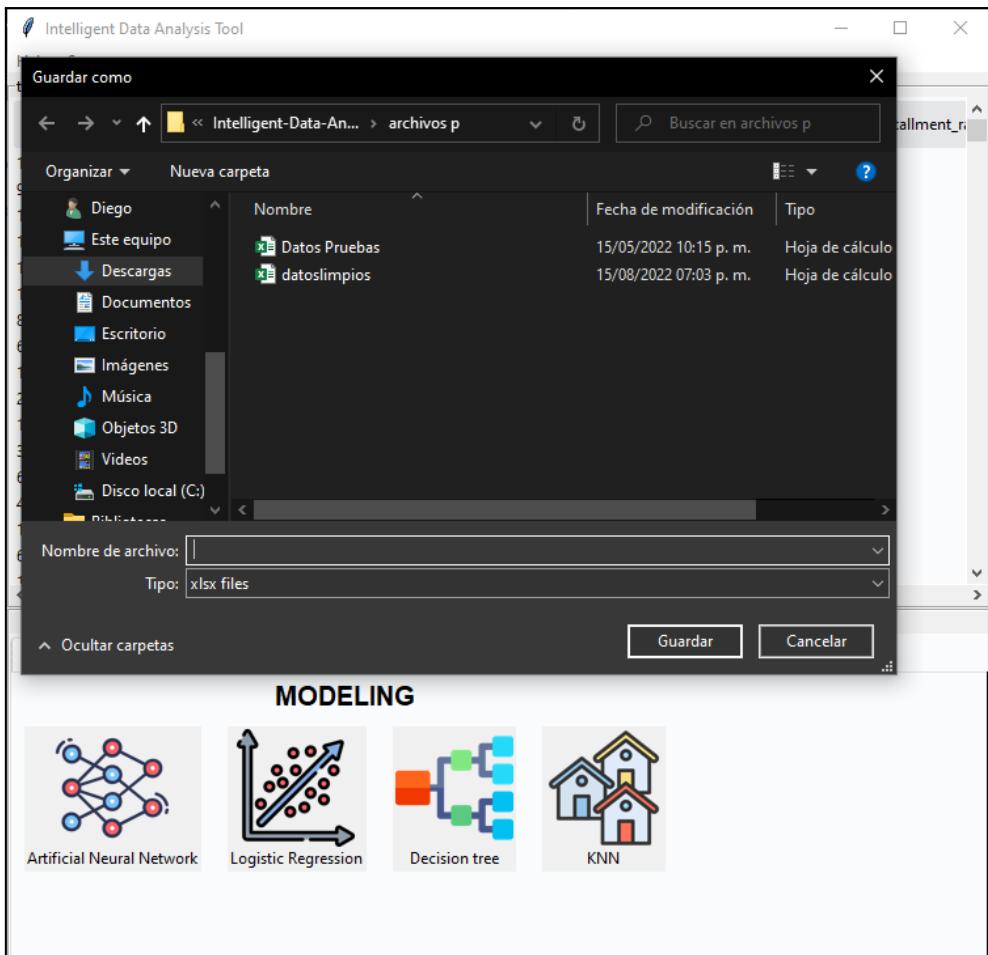


Figura 4.1: Guardando dataset.

Al seleccionar la opción “Save As ” desplegará un menú, en el que se podrá elegir el nombre con el que se guardará el proyecto, como lo podemos observar en “Figura 4.1”

A	B	duration	amount	employment_duration	installment_rate
duration	amount	18.0	1049.0	2.0	4.0
18	1049	9.0	2799.0	3.0	2.0
9	2799	12.0	841.0	4.0	2.0
12	841	12.0	2122.0	3.0	3.0
12	2122	12.0	2171.0	3.0	4.0
12	2171	10.0	2241.0	2.0	1.0
10	2241	8.0	3398.0	4.0	1.0
8	3398	6.0	1361.0	2.0	2.0
6	1361

Figura 4.2: Comprobación de datos guardados.

Como podemos ver en la figura 4.2 podemos rectificar el correcto funcionamiento de la aplicación al guardar datos.

Modelado

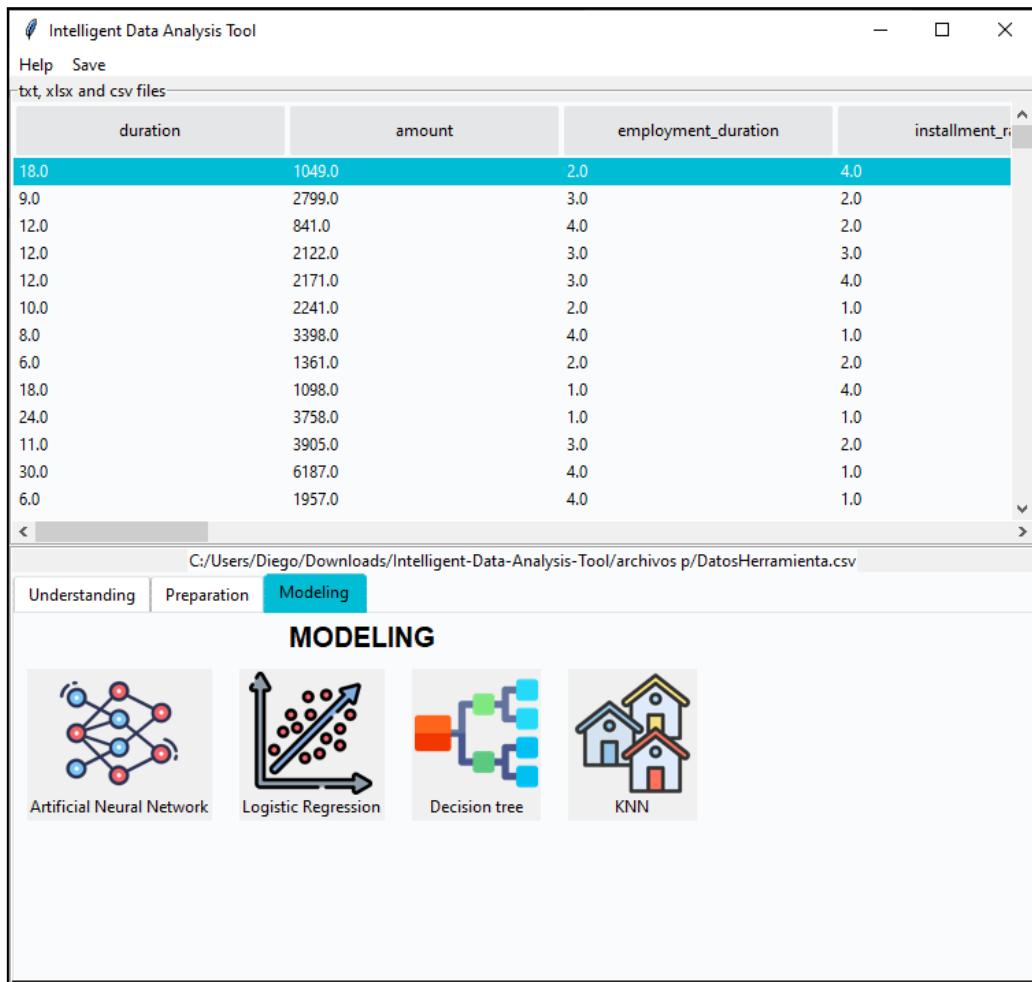


Figura 5.0: Modelado de datos.

Para el modelado de los datos se cuenta con cuatro diferentes modelos, como se muestra en la "Figura 5.0"

Los modelos que puede generar la herramienta se centran en problemas de clasificación, y dichos modelos son:

1. Artificial Neural Network.
2. Linear Regression.
3. Decision tree.
4. KNN.

Proceso para generar un modelo

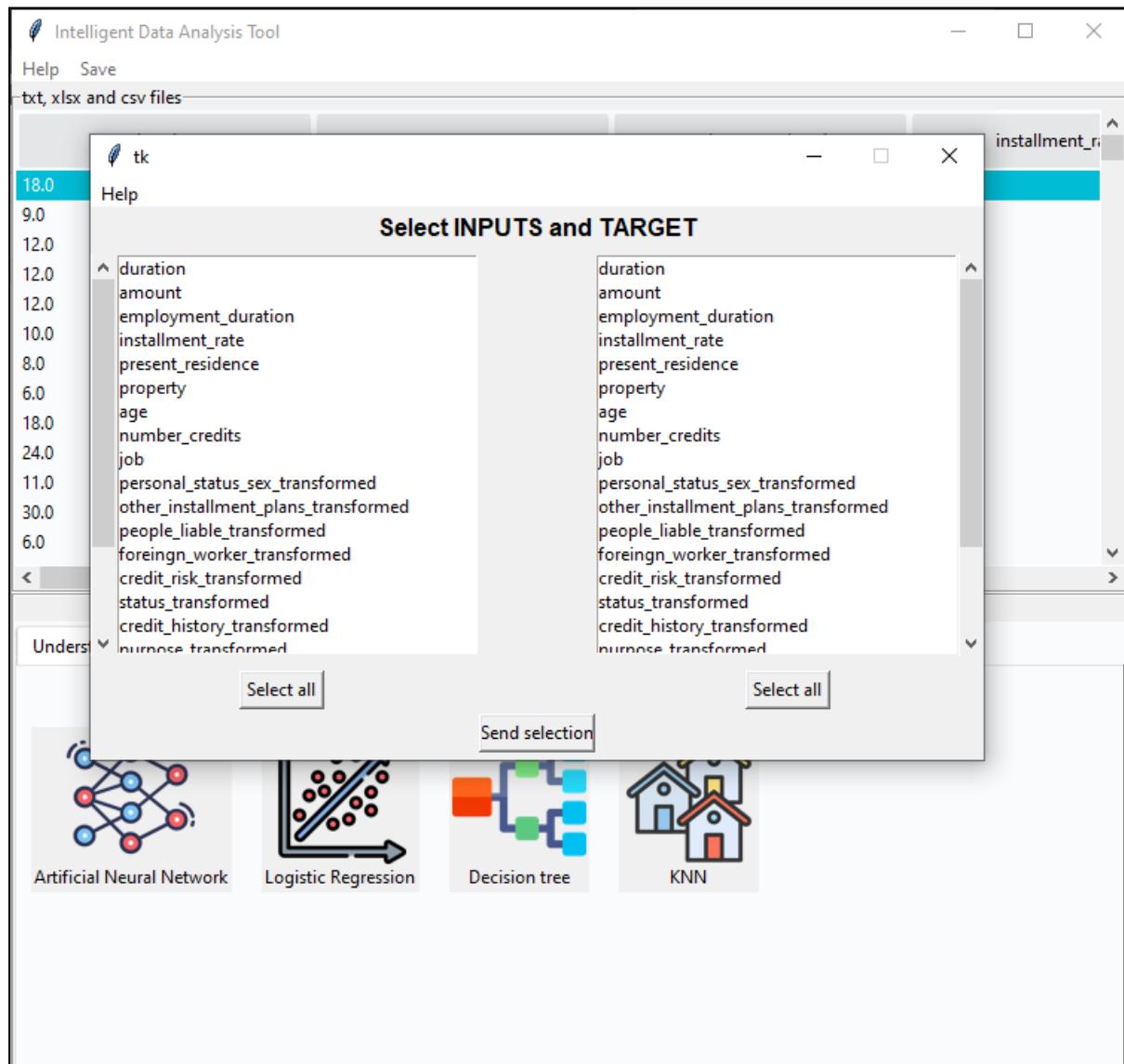


Figura 5.1: Entradas y Objetivos sobre modelado.

Al seleccionar una de las opciones contenidas en modelado, se desplegará una ventana, la cual esperará entradas y sus respectivos objetivos. Una vez se haga la selección como se muestra en la “Figura 5.1”, al terminar de hacer selección respectiva presionamos “Send selection”

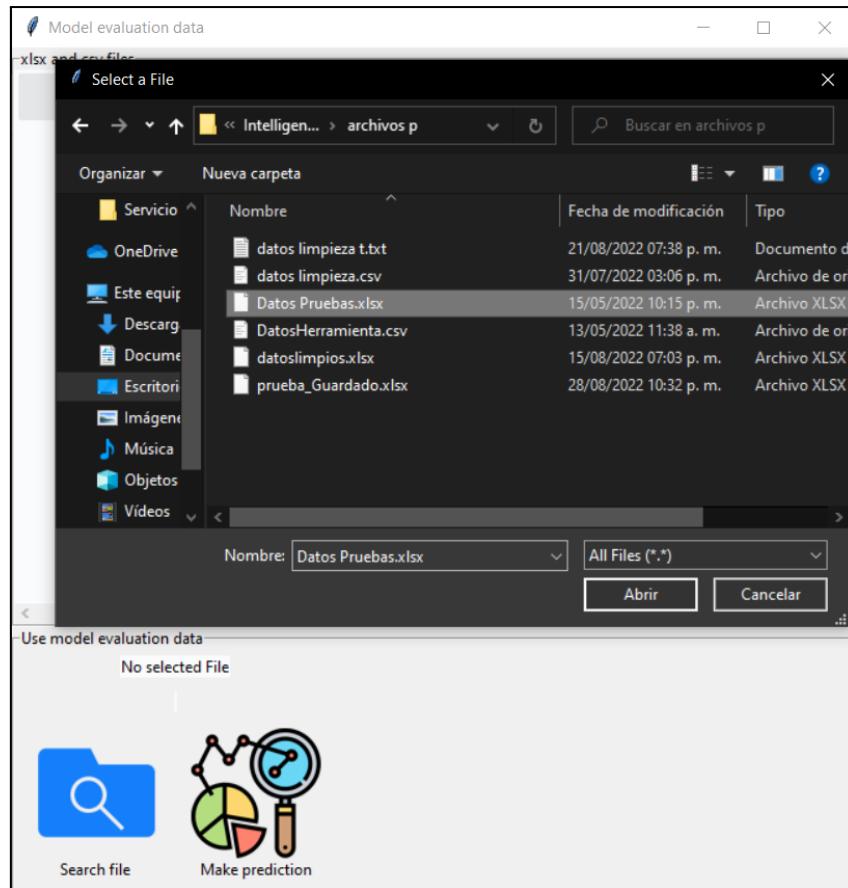


Figura 5.2: Cargando datos de evaluación del modelo.

Al terminar con la selección de entradas y objetivos, procedemos a seleccionar “Search file” para cargar los datos de la evaluación del modelo como se muestra en la “Figura 5.2”.

duration	amount	employment_duration
12	691	5
42	4370	4
36	2746	3
24	4110	5
18	2462	3
12	1282	2
12	2969	2
48	4695	4

Datos de prueba cargados.

Figura 5.3: Datos de prueba cargados.

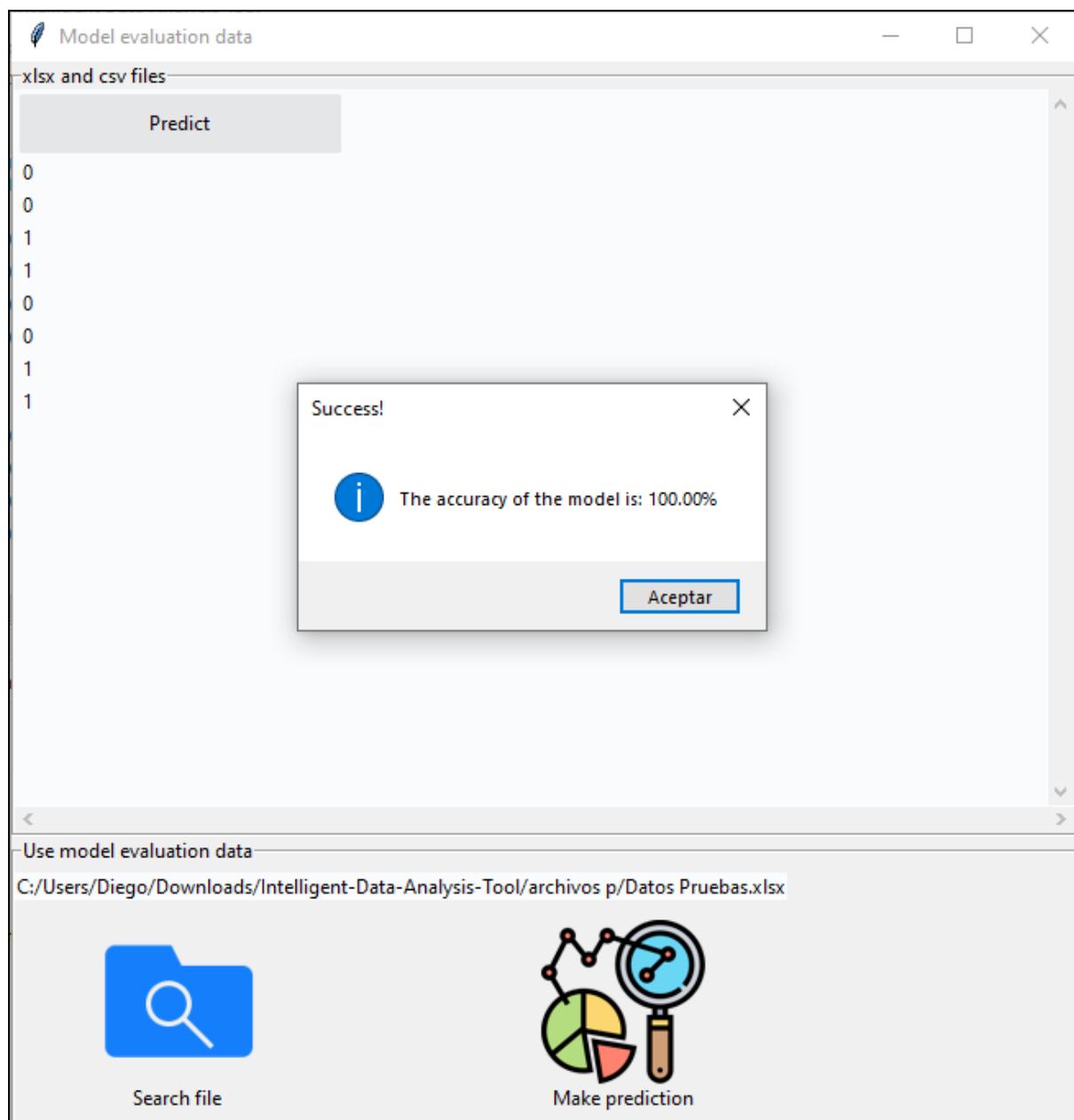


Figura 5.3: Predicción realizada.

Una vez cargado el archivo de prueba, seleccionamos “Make prediction” para generar el modelo y evaluarlo con los datos seleccionados. Al terminar dicha ejecución se desplegará una ventana en la cual se indicará la bondad del modelo. Esto lo podemos observar en la “Figura 5.3”.

Difusión - Página web

Maquetación de la página de difusión, donde vendrá contenido el repositorio donde está el código de la herramienta, repositorio UCI Machine Learning y documentación de la metodología CRISP DM.

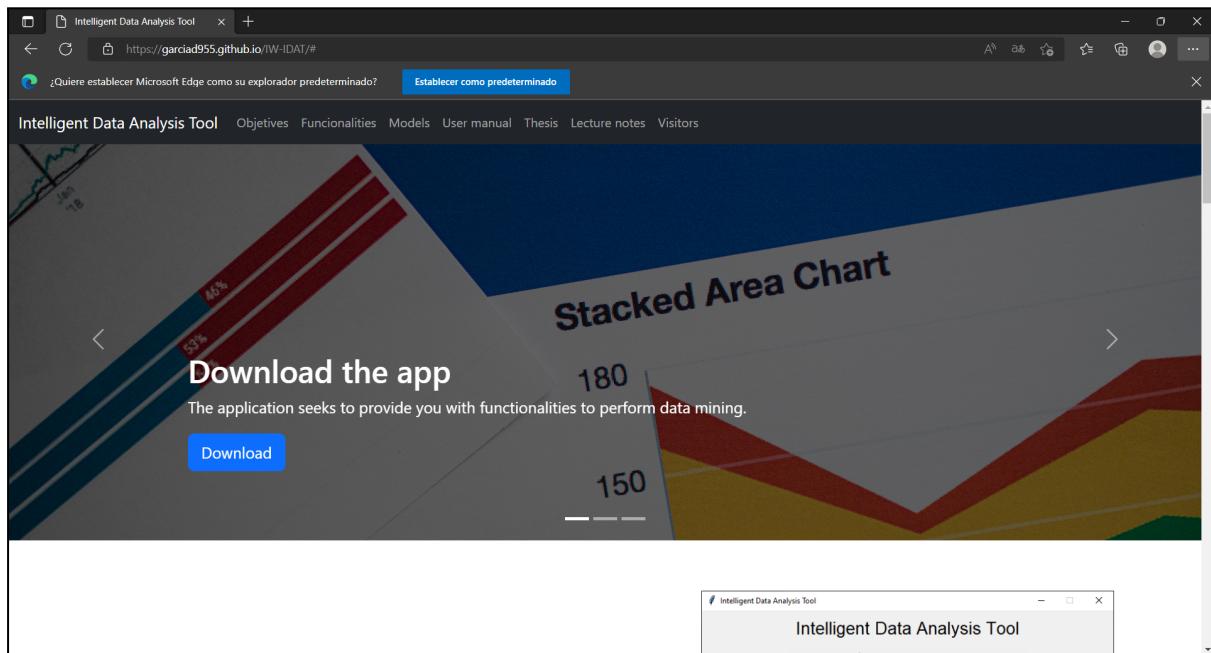


Figura 6.0: Vista descarga aplicación.

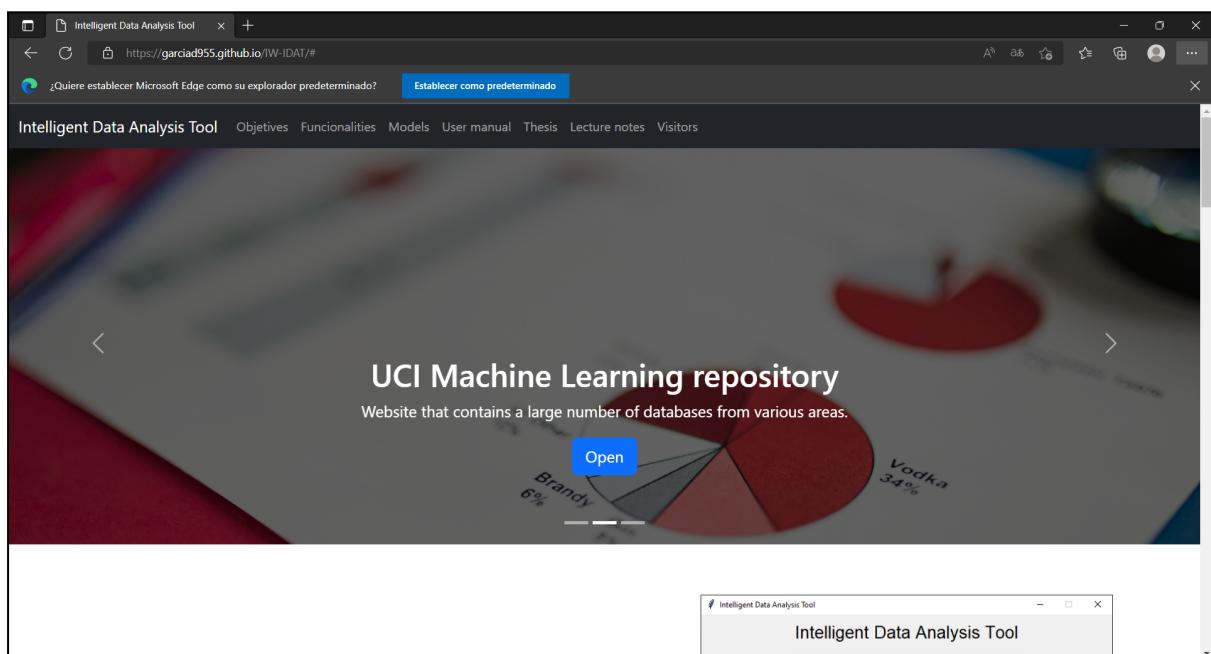


Figura 6.1: Vista repositorio UCI Machine Learning.

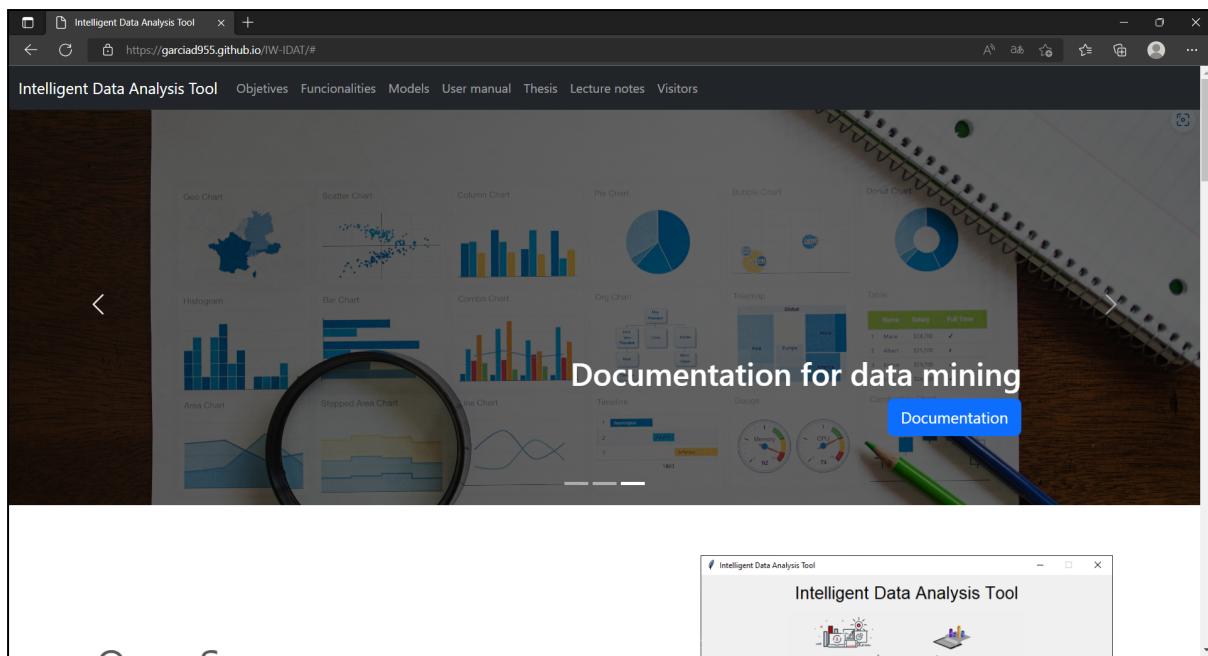


Figura 6.2: Vista documentación metodología CRISP DM.

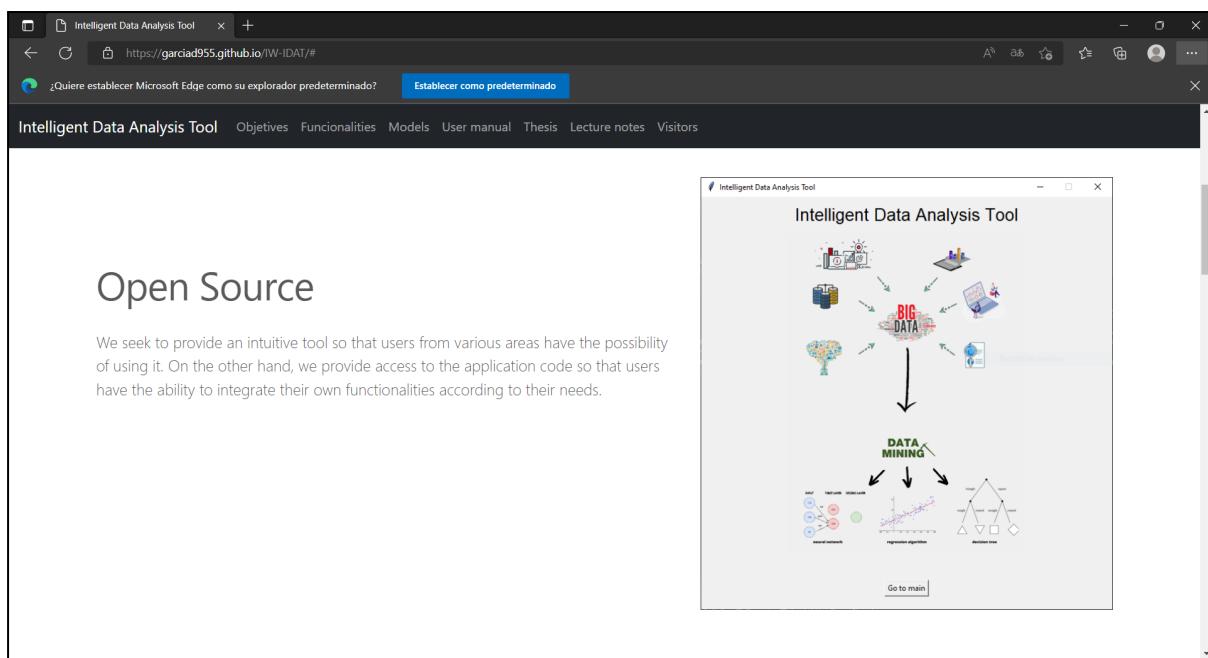


Figura 6.2: Vista objetivos.

Intelligent Data Analysis Tool x +

https://garciad955.github.io/IW-IDAT/#modelos Establecer como predeterminado

Intelligent Data Analysis Tool Objetives Funcionalities Models User manual Thesis Lecture notes Visitors

Modeling techniques

The tool has several types of modeling, such as:

- Artificial Neural Network
- Decision trees
- Regression algorithms
- K nearest neighbors algorithm

Designed for the user to determine the most appropriate model according to the problem he is working on.

Red Neuronal Artificial

The diagram illustrates an Artificial Neural Network (Red Neuronal Artificial). It consists of three layers: Input layer, Hidden layer, and Outputlayer. The Input layer has five nodes labeled Data1, Data2, Data3, Data4, and Data5. The Hidden layer has four nodes. The Outputlayer has one node. Every node in the Input layer is connected to every node in the Hidden layer, and every node in the Hidden layer is connected to the single node in the Outputlayer. The connections are represented by black lines.

Figura 6.2: Vista modelado.

Intelligent Data Analysis Tool x +

https://garciad955.github.io/IW-IDAT/#user-manual

Intelligent Data Analysis Tool Objetives Funcionalities Models User manual Thesis Lecture notes Visitors

User manual

The screenshot shows the 'User manual' section of the Intelligent Data Analysis Tool. It displays a single page from a PDF titled 'Herramienta Intelligent Data Analysis Tool'. The page contains text explaining the purpose and features of the tool, mentioning its ability to handle datasets and implement machine learning models. The text is in Spanish.

Figura 6.2: Vista manual de usuario.

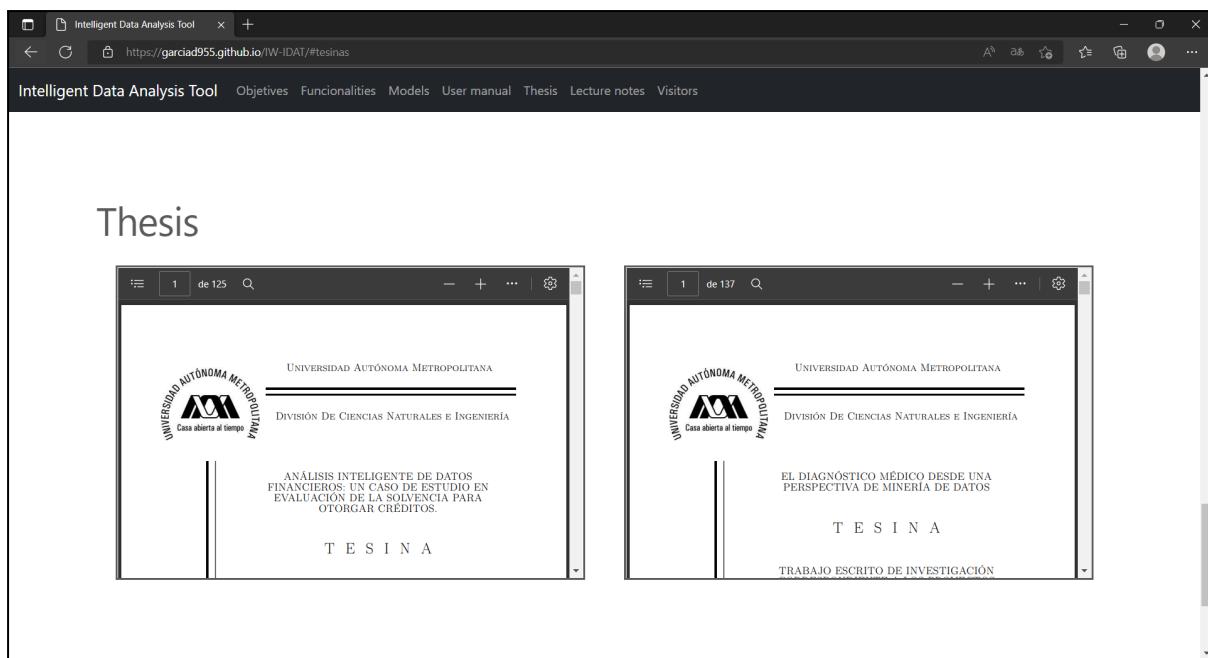


Figura 6.2: Vista tesinas.

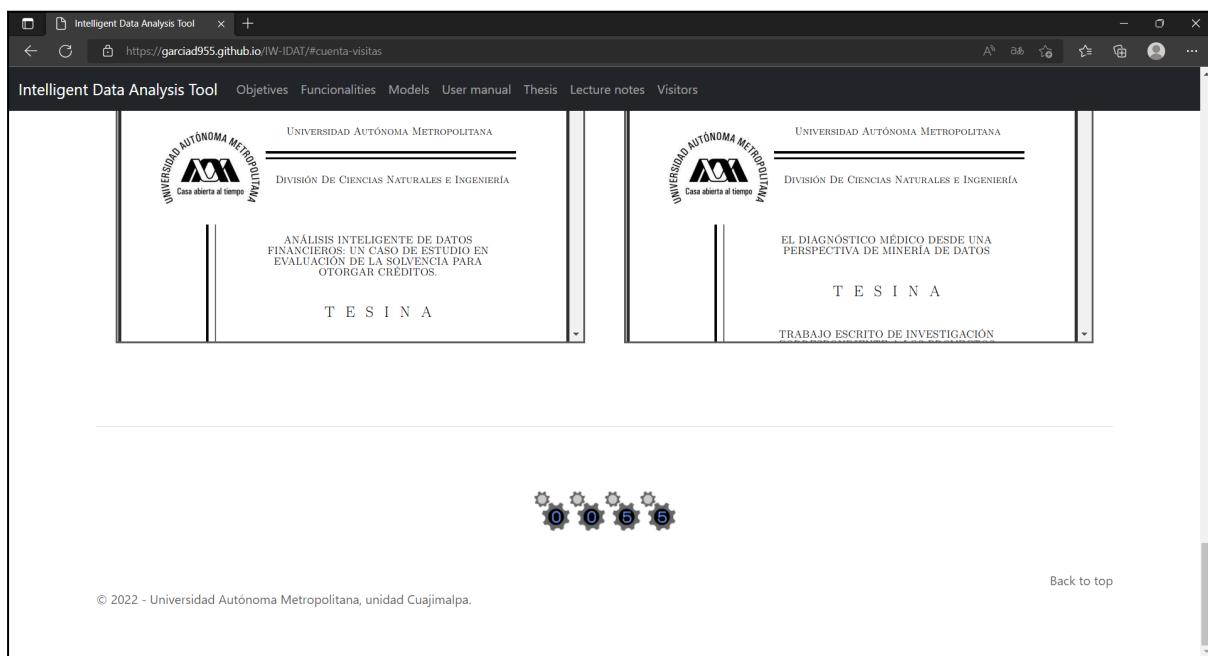


Figura 6.2: Vista contador de visitas.