



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE INGENIERÍA

**Proyecto Final  
Documentación GUI**

REALIZADA POR  
**Gómez Luna, Alejandro**  
**316034946**

PROFESOR  
**Dr. Molero Castillo, Guillermo Gilberto**

ASIGNATURA  
**Inteligencia Artificial**

**GRUPO No. 3**

FECHA  
**8/Diciembre/2021**

## Objetivos

- Poder crear una interfaz gráfica que le permita a un usuario con conocimiento en el campo de la Inteligencia Artificial interactuar con ella y poder utilizar los distintos algoritmos revisados en clase sobre un conjunto de datos que se encuentren en algún archivo, el cual será seleccionado por el usuario.
- Poder crear una vista dentro de la interfaz donde un usuario no experto relacionada con algún área sobre la que se quiera realizar pronósticos tendrá acceso para poder ingresar datos y con base en ellos obtener una predicción.

## Requerimientos

### ■ Funcionales

- *Requerimientos generales*
  - La interfaz debe de contar con un apartado para que el usuario pueda subir los archivos que contengan los datos que se quieran analizar.
  - Se debe de tener un apartado en donde se puedan visualizar todos los datos que se ingresaron a través del archivo.
  - El usuario debe de poder seleccionar el algoritmo que desee aplicar al conjunto de datos introducido.
  - Se debe de tener un apartado en donde se puedan emplear los distintos algoritmos enfocados a un caso de estudio en particular.
- *Implementación de reglas de asociación a través del algoritmo apriori*
  - Se le debe mostrar al usuario una tabla y/o gráfico en donde se puedan visualizar la frecuencia de cada instancia en el conjunto de datos proporcionado, para que el usuario pueda determinar los elementos que más apariciones tienen.
  - Ofrecer un apartado para que el usuario pueda ingresar los valores de los parámetros lift, soporte y confianza para que así se puedan generar las respectivas reglas de asociación tomando como base dichos parámetros ingresados. Para cada regla de asociación se debe de mostrar los elementos que la conforman, y los parámetros de lift, soporte y confianza de dicha regla.
  - Mostrar un respectivo apartado para que el usuario pueda introducir sus conclusiones respecto a las reglas de asociación obtenidas y, una vez que haya finalizado de introducir dichas, guardarlas en un respectivo archivo de texto.
- *Implementación de métricas de distancia*
  - El usuario debe de poder seleccionar una de las siguientes métricas de distancia: Euclidiana, Chebyshev, Manhattan y Minkowsky (Para esta métrica también se debe de especificar el valor de la lambda). Cada una de estas métricas se calculará respecto a los datos introducidos por el usuario.
  - También se le debe de ofrecer la opción al usuario para que seleccione si quiere calcular las métricas de distancia sobre todo el conjunto de datos, solo una porción o solamente para un par de elementos.
- *Implementación de clusterización*
  - Se le debe de mostrar al usuario una matriz de correlaciones y/o mapa de calor para que pueda visualizar las relaciones que existen entre las variables de sus datos y así se pueda realizar una selección de características.
  - En la selección de características el usuario podrá escoger aquellas variables que desea tomar en cuenta o no para el proceso de clusterización.

- En la clusterización jerárquica se debe de generar un dendograma para que el usuario sepa el número de clústers que deberá de seleccionar para aplicar el algoritmo ascendente jerárquico.
- En la clusterización particional se debe de generar un gráfico poligonal para que el usuario pueda aplicar el método del codo y determinar el valor de k para aplicar el algoritmo de k-means. Como muchas veces este método es heurístico es bastante recomendable agregar un localizador del codo para que sea más explícito el valor de k que debe de utilizar el usuario.

Para ambos tipos de clusterización, una vez que el usuario haya ingresado el número de clústers, se debe de mostrar los clústeres generados con sus respectivos centroides y cantidad de elementos que los conforman. Asimismo, también se debe de incluir un apartado para que el usuario genere sus respectivas conclusiones respecto a los clústers generados y estos se almacenen en un archivo de texto.

- *Implementación de clasificación con regresión logística*

- Se le debe de mostrar al usuario una matriz de correlaciones y/o mapa de calor para que pueda visualizar las relaciones que existen entre las variables de sus datos y así se pueda realizar una selección de características.
- En la selección de características el usuario podrá escoger aquellas variables que desea tomar en cuenta o no para el proceso de clasificación.
- Además de la selección de variables el usuario debe de poder escoger la variable sobre la cual querrá realizar el proceso de predicción, mientras que todas las demás variables no descartadas en el proceso de selección serán utilizadas para realizar la predicción. Como parte de estas configuraciones previas el usuario también debe de poder escoger el porcentaje de datos que utilizará para la fase de pruebas, dejando el resto para la fase de entrenamiento del modelo de regresión logística.
- Una vez que se hayan ingresado correctamente todos estos parámetros, se mostrará en pantalla la matriz de confusión generada y el modelo que se generó, así como sus respectivos parámetros de exactitud, precisión, tasa de error, sensibilidad, especificidad y la eficiencia.
- Después de que se haya generado el modelo para realizar predicciones entonces existirá otra vista con la cual pueda interactuar un usuario experto en el área de los datos que se están analizado, y la cual será utilizada para que, con base en las variables tomadas en cuenta para la predicción, estas se ingresen a través de dicha vista y sean tomadas en cuenta para realizar la predicción de la variable clase escogida previamente. Aunado al resultado de la predicción se debe de mostrar la eficiencia del modelo con el cual se realizó dicha predicción.

- *Implementación de pronóstico y clasificación con árboles de decisión*

- Se le debe de mostrar al usuario una matriz de correlaciones y/o mapa de calor para que pueda visualizar las relaciones que existen entre las variables de sus datos y así se pueda realizar una selección de características.
- En la selección de características el usuario podrá escoger aquellas variables que desea tomar en cuenta o no para el proceso de clasificación.
- Además de la selección de variables el usuario debe de poder escoger la variable sobre la cual querrá realizar el proceso de predicción, mientras que todas las demás variables no descartadas en el proceso de selección serán utilizadas para realizar la predicción. Como parte de estas configuraciones previas el usuario también debe de poder escoger el porcentaje de datos que utilizará para la fase de pruebas, dejando el resto para la fase de entrenamiento del modelo de regresión logística.

También en las configuraciones previas el usuario debe de poder ajustar parámetros tales como el nivel de profundidad máximo del árbol que se vaya a generar, la cantidad mínima de elementos que deben de existir en los nodos intermedio y la cantidad mínima de elementos que deben de existir en los nodos hoja.

Por último, se debe de poder escoger entre crear un árbol de decisión para pronóstico o un árbol de decisión para clasificación.

- Una vez que se hayan ingresado correctamente todos estos parámetros, el resultado que se mostrará en pantalla podrá ser
    - ◇ Para pronóstico: Los parámetros del modelo generado, tales como MAE, MSE, RMSE y Score, para que se puedan caracterizar las predicciones que realiza nuestro modelo.
    - ◇ Para clasificación: La matriz de confusión generada, así como sus respectivos parámetros de exactitud, precisión, tasa de error, sensibilidad, especificidad y la eficiencia.
  - En ambos casos se debe de mostrar una matriz donde se muestre las variables en conjunto con su porcentaje de importancia ordenados de mayor a menor, para analizar aquellas variables que en el proceso de predicción tienen mayor relevancia y aquellas que no.
- Al igual que esta matriz debe de ser posible visualizar el árbol de decisión que se generó, ya sea de manera detallada o un panorama general, en formato de imagen, así como una representación del árbol de forma detallada en formato de texto, para que el usuario pueda analizar de manera más minuciosa los nodos que conforman al árbol y las características de dichos nodos.
- Después de que se haya generado el modelo para realizar predicciones entonces existirá otra vista con la cual pueda interactuar un usuario experto en el área de los datos que se están analizado, y la cual será utilizada para que, con base en las variables tomadas en cuenta para la predicción, estas se ingresen a través de dicha vista y sean tomadas en cuenta para realizar la predicción de la variable clase escogida previamente. Aunado al resultado de la predicción se debe de mostrar la eficiencia del modelo con el cual se realizó dicha predicción.

#### ■ No Funcionales

- Ofrecer una interfaz gráfica lo más intuitiva posible para que el usuario pueda interactuar con ella de manera dinámica y sencilla.
- Mostrar el contenido de una manera que el usuario pueda emplear y configurar los algoritmos de una manera gráfica, además de mostrar los resultados correspondientes de manera adecuada.
- La interfaz cuenta con la menor cantidad de limitantes para que el usuario pueda utilizarla sin la necesidad de que realice un laborioso proceso de instalación.

### Metodología de trabajo

1. Primero se realizó un análisis de las distintas tecnologías y herramientas que se pudiera utilizar para el desarrollo del proyecto. De cada una de dichas herramientas y tecnologías previstas se realizó un análisis de las ventajas como desventajas que podría ofrecer, así como el tiempo requerido para poder aprenderlas. Durante este proceso de análisis se fueron descartando opciones hasta finalmente obtener el conjunto de herramientas a utilizar.
2. Posteriormente se realizó el siguiente plan de trabajo semanal

Día de la semana	<i>Martes</i>	<i>Jueves</i>	<i>Sábado</i>	<i>Domingo</i>
ACTIVIDADES	Análisis de requerimientos para el proyecto		Trabajar en el proyecto	

Figura 1: Plan de trabajo semanal

Donde las actividades se encuentran conformadas por las siguientes etapas

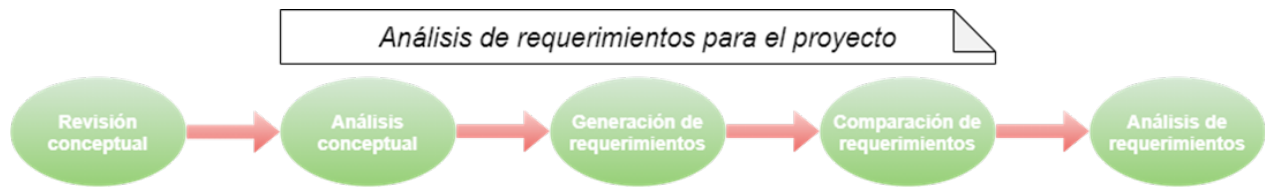


Figura 2: Diagrama para la actividad de análisis de requerimientos

- *Revisión conceptual*: Adquirir nuevos conocimientos respecto a conceptos y algoritmos relacionados con el campo de la IA.
- *Análisis conceptual*: Analizar cada uno de estos nuevos conceptos y algoritmos para poder entender su importancia y/o funcionamiento.
- *Generación de requerimientos*: Con base en este proceso de análisis se generará una lista de requerimientos que deberán de ser implementados en el proyecto.
- *Comparación de requerimientos*: Se verifica si previamente ya se tenía una lista de requerimientos y se revisa si esta lista de requerimientos previa necesita ser actualizada, modificada o se mantiene intacta.
- *Análisis de requerimientos*: Realizar un panorama general de trabajo para poder implementar la lista de requerimientos solicitados.



Figura 3: Diagrama para la actividad de trabajar en el proyecto

- *Generar plan de trabajo*: Conforme al análisis de requerimientos se generará un plan de trabajo a seguir para poder cumplir con lo que se solicita.
- *Análisis previo a codificación*: Es la etapa en donde se realiza un proceso de análisis para poder saber cómo realizar la correspondiente transformación de requerimientos a un lenguaje de programación. Para este proceso se generan distintos elementos que faciliten el proceso de análisis, como lluvias de ideas, esquemas, pseudocódigos, etc.
- *Codificación*: Ya se realiza la codificación de requerimientos utilizando las herramientas y lenguajes de programación escogidos.
- *Testing*: Se realizan algunas pruebas para observar el correcto funcionamiento del sistema. En caso de que se presenten errores, se vuelve al proceso de análisis previo a codificación para darle solución a dichos errores y, una vez que las pruebas estén libres de errores, entonces se puede pasar a la siguiente etapa.
- *Refinamiento*: Una vez que se ha asegurado que el sistema funciona correctamente, se realiza un proceso de mejoramiento en donde a la interfaz se le añaden mejores elementos gráficos, cambio de colores, etc.

## Herramientas utilizadas

- *HTML5, CSS3, Bootstrap 5, JavaScript y JQuery*: Estas herramientas se utilizaron para poder realizar toda la parte gráfica de la interfaz, es decir, el frontend. Con HTML5, CSS3 y Bootstrap 5 se puede personalizar una página web de manera bastante sencilla, pues con HTML5 se pueden crear distintos elementos como selectores, cajas de texto, etc, y en conjunto CSS3 con Bootstrap 5 nos permitirán personalizar dichos elementos para que tengan una mejor apariencia. Por otro lado, JavaScript y JQuery nos permitirán poder manipular dichos objetos de manera dinámica, para poder obtener la información ingresada por el usuario, cargar nuevos elementos en la página web después de cierta acción, etc.
- *Chart.js y File.js*: Son bibliotecas desarrollados en JavaScript que permiten generar gráficas de barras, pastel, etc, así como la manipulación de archivos, respectivamente.
- *Python*: Este lenguaje de programación es de alto nivel y multiparadigma, lo que permite que se puedan realizar muchas operaciones de manera sencilla. Para este caso, servirá como la parte del backend de la aplicación, a la cual se le pasará los datos ingresados por el usuario y se encargará de aplicar los algoritmos necesarios para generar información que será presentada al usuario.
- *Pyhton Eel*: Es una biblioteca de Python que permite interconectar la parte del frontend con la parte del backend, para que exista esta constante comunicación con lo que ingresa el usuario, el procesamiento de dichos datos y derivado de ello presentar la información necesaria en pantalla.

## Descripción

Se va a tener una primer pantalla con los siguientes elementos

1. Este primer elemento es un botón que nos permitirá regresar nuevamente a esta página inicial.
2. Este segundo elemento es un botón que nos permitirá visualizar los elementos que se encuentren en el archivo subido por el usuario. Si dicho archivo tiene más de 100 elementos, solamente se mostrarán los primero 50 y últimos 50 elementos.
3. Este tercer elemento es un botón que nos permitirá implementar una interfaz que será usada por algún usuario final experto en el tema sobre el cual se esté realizando una predicción conforme a los datos introducidos.
4. Este cuarto elemento es una pestaña despegable, la cual contiene en cada renglón cada uno de los algoritmos que se pueden implementar en esta interfaz gráfica.
5. Este quinto elemento corresponde al apartado para que el usuario pueda subir el archivo al que le desea aplicar algún análisis. Cabe resaltar que todos los archivos se deben encontrar en el directorio de trabajo del proyecto para que puedan ser encontrados y cargados correctamente. Para seleccionar el archivo se abrirá el explorador de archivos, desde donde se podrá seleccionar el que se desea. Solamente se puede ingresar uno y este debe de tener una terminación .csv, .xls o .xlsx. En este apartado también se especifica si se desea tomar en cuenta que la primera fila corresponde a la cabecera que define el nombre de las columnas o no.  
Por último, ya que se ha seleccionado el archivo, se habilitan los botones de SUBIR ARCHIVO y LIMPIAR, con los cuales se podrá subir el archivo y eliminar la selección de archivo, respectivamente. Una vez que se suba el archivo, los elementos en la barra de navegación estarán disponibles para ser usados.

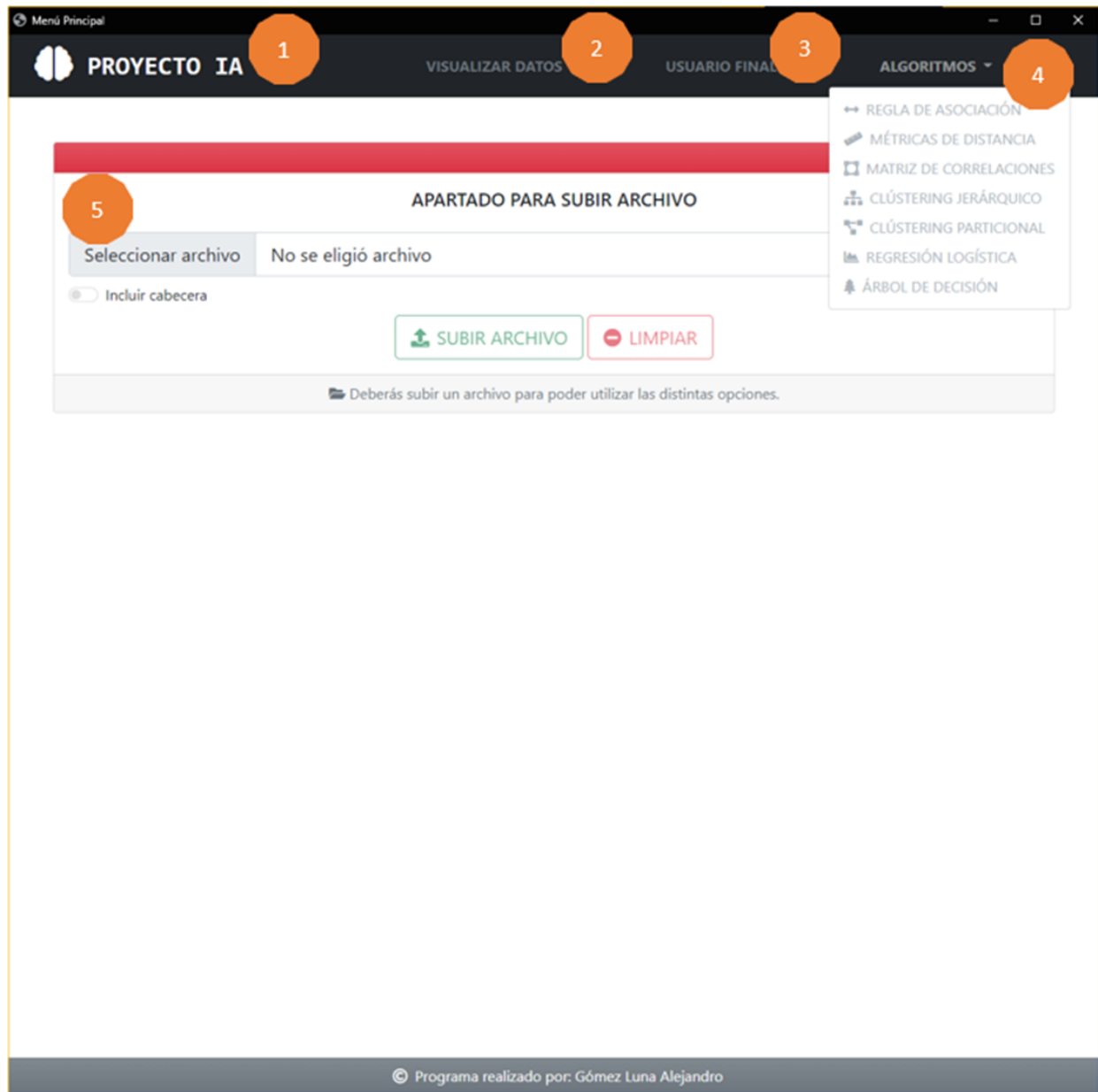


Figura 4: Pantalla principal de la GUI

A partir de esta primer pantalla se van a ir desglosando las demás vistas con las cuales cuenta la interfaz, en donde la mayoría corresponde a un algoritmo, y las demás son parte de diversos análisis previos utilizados para obtener características de los datos introducidos

## 1. Desglose de la pestaña VISUALIZAR DATOS

Mostrar  registros

Buscar:

#	IDNumber	Diagnosis	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints
0	P-842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471
1	P-842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017
2	P-84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279
3	P-84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052
4	P-84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043
5	P-843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089
6	P-844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074
7	P-84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985
8	P-844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353

Mostrando registros del 1 al 10 de un total de 569 registros

Anterior **1** 2 3 4 5 ... 57 Siguiente

© Programa realizado por: Gómez Luna Alejandro

Figura 5: Vista para visualizar los datos del archivo

En esta pestaña se permitirá visualizar los datos contenidos en el archivo.

La primera fila de encabezado podrá o no contener los nombres de las columnas, dependiendo de la selección realizada previamente en el apartado de subir archivo.

Con esta tabla nosotros podremos utilizar diversas funcionalidades como

- Mostrar una cierta cantidad de registros en la página, para poder observar mayor cantidad de registros.
- Buscar algún elemento que nosotros deseemos, solamente ingresando el valor.
- Consultar registros en alguna otra página mediante la barra de navegación que tiene los botones de anterior, siguiente y el número de página en el que nos encontramos. Asimismo, se muestra la cantidad de registros totales y cuántos se están mostrando en este momento.



- Ordenar los registros de alguna columna de mayor a menor, para poder hacer un mejor análisis de los datos

## 2. Desglose del apartado REGLAS DE ASOCIACIÓN

Regla Asociación

PROYECTO IA

VISUALIZAR DATOS USUARIO FINAL ALGORITMOS

Mostrar 10 registros

Buscar:

#	Item	Frecuencia	Porcentaje
119	Ninja Turtles	1786	0.0609119744
118	Get Out	1346	0.0459056649
117	Tomb Rider	1305	0.0445073497
116	Hotel Transylvania	1280	0.0436547185
115	Coco	1229	0.0419153508
114	John Wick	989	0.0337300911
113	Moana	971	0.0331161966
112	Jumanji	736	0.0251014631
111	Intern	714	0.0243511476
110	Spotlight	712	0.0242829371

Mostrando registros del 1 al 10 de un total de 119 registros

Anterior 1 2 3 4 5 ... 12 Siguiete

Figura 6: Primera parte de la vista para el algoritmo de reglas de asociación

En esta pestaña se permite implementar el algoritmo de apriori para generar distintas reglas de asociación. Se empieza por mostrar una matriz donde se observa la frecuencia de veces que aparece un elemento en todo el conjunto de datos proporcionados, así como el porcentaje total que representa del total de datos.

Esta tiene las mismas funcionalidades que la tabla de la vista para visualizar datos.

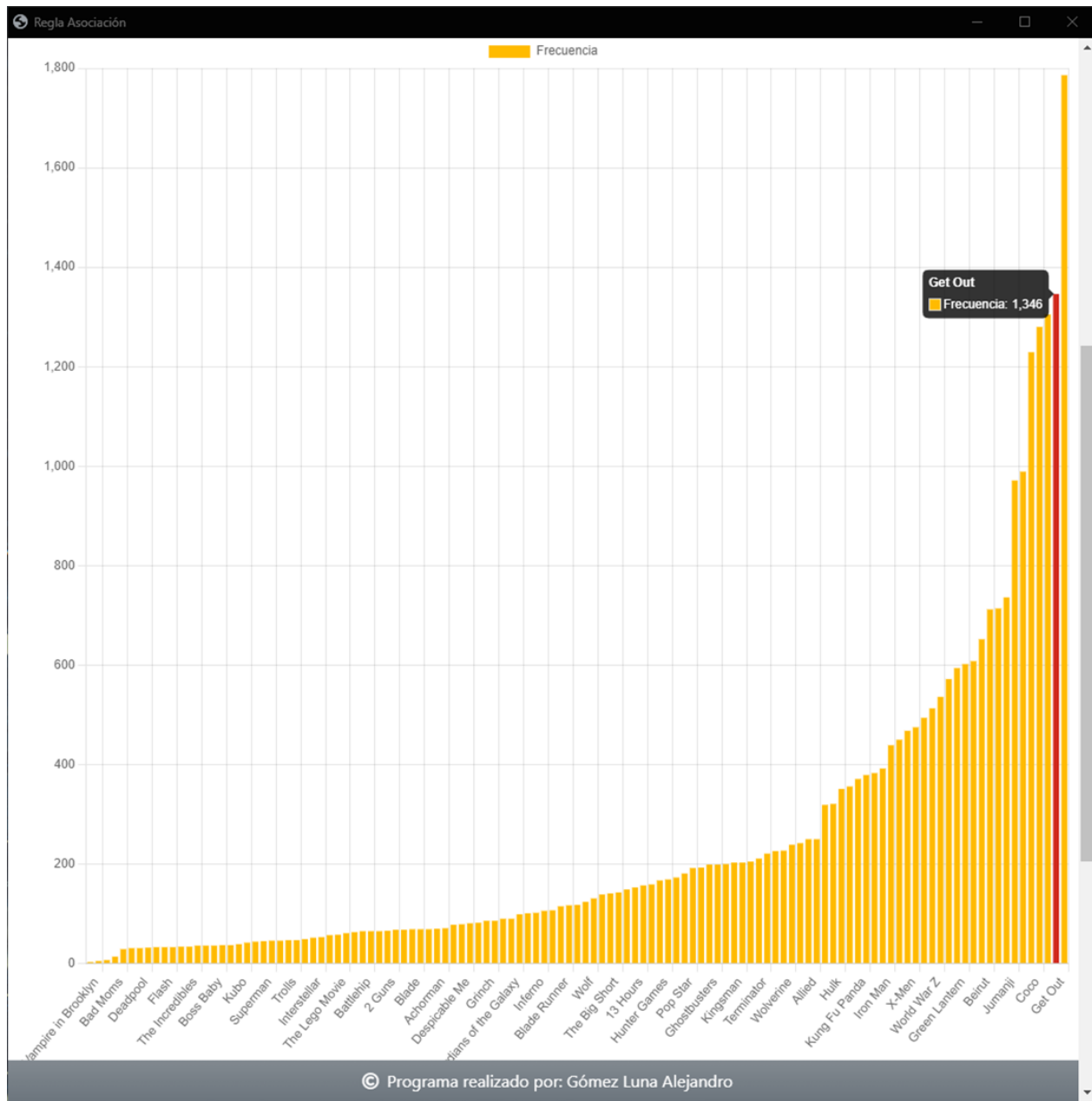


Figura 7: Segunda parte de la vista para el algoritmo de reglas de asociación

Posteriormente se muestra una gráfica de barras con base en la tabla anterior, con la cual se permite visualizar de mejor manera la frecuencia que tiene cada uno de los elementos en el conjunto de datos proporcionado. Si se quiere visualizar con más detalle la frecuencia de alguno de los elementos se necesita posicionar el cursor sobre la barra correspondiente y esta se sobresaltará de color rojo, mostrando en la punta el nombre del elemento junto con su frecuencia.

**Ingresar Parámetros**

**Soporte**   
Solamente ingresar el numero como en el ejemplo. El valor debe estar en formato decimal y deberá ser menor a 1.

**Confianza**   
Solamente ingresar el numero como en el ejemplo. El valor debe estar en formato decimal y deberá ser menor a 1.

**Lift**   
Solamente ingresar el numero como en el ejemplo.

Regla #1 ▼

Regla #2 ▲

Regla de asociacion: Black Panther, Ninja Turtles

Soporte: 0.027613941018766755

Confianza: 0.33881578947368424

Lift: 1.4152104084399129

Regla #3 ▼

Regla #4 ▼

Regla #5 ▼

Regla #6 ▼

© Programa realizado por: Gómez Luna Alejandro

Figura 8: Tercera parte de la vista para el algoritmo de reglas de asociación

Finalmente se tiene el apartado para que se puedan ingresar los parámetros deseados de soporte, confianza y lift para generar distintas reglas de asociación. Una vez ingresados estos datos se desplegará una lista de secciones despegables. Cada sección corresponde a una regla de asociación, la cual para visualizarla se da click sobre la sección deseada y se mostrará la regla de asociación, soporte, confianza y lift, así como un área de texto para generar la respectiva conclusión sobre dicha regla.

Hasta la parte inferior se presenta un botón para guardar dichas conclusiones. Al hacer click sobre dicho botón se desplegará el explorador de archivos para que nosotros almacenemos las conclusiones generadas en algún directorio de nuestra computadora.

### 3. Desglose del apartado MÉTRICAS DE DISTANCIA

**Ingresar Parámetros**

Seleccionar Métrica:

Lambda:   
Parametro para Minkowski únicamente. Solamente ingresar el numero como en el ejemplo. Se permiten decimales.

Valores: ☐ Calcular toda la matriz de distancias  
☒ Calcular un subconjunto de distancias  
☐ Calcular una pareja de valores

Indices:   
Para obtener un suconjunto indicar los indices que se incluiran separados por un -

Mostrar  registros

Buscar:

#	T1	10	T1	11	T1	12	T1	13	T1	14	T1	15	T1	16	T1	17	T1	18	T1	19	T1	20	T1	21
10	0		153234.541		66453.217		250454.087		64488.746		274420.688		218862.212		34398.312		36354.078		297939.431		124228.55		3	
11	153234.541	0		102008.399		97594.429		216122.022		122390.22		66474.492		169250.698		172862.903		145001.112		29173.817		1		
12	66453.217	102008.399	0		191948.474		124977.298		223208.741		159220.364		69081.085		73036.875		238496.96		75110.671		4			
13	250454.087	97594.429	191948.474	0		312957.064		35194.662		32827.066		260523.053		264267.868		47507.675		126302.809		2				
14	64488.746	216122.022	124977.298	312957.064	0		337736.347		281101.579		58798.6		54969.102		360447.435		187122.736		8					
15	274420.688	122390.22	223208.741	35194.662	337736.347	0		66425.668		291084.344		294721.429		30185.397		150873.915		2						
16	218862.212	66474.492	159220.364	32827.066	281101.579	66425.668	0		228007.559		231829.888		79578.052		94985.011		2							
17	34398.312	169250.698	69081.085	260523.053	58798.6	291084.344	228007.559	0		4465.644		307429.249		141421.438		2								
18	36354.078	172862.903	73036.875	264267.868	54969.102	294721.429	231829.888	4465.644	0		311257.274		144920.782		2									
19	297939.431	145001.112	238496.96	47507.675	360447.435	30185.397	79578.052	307429.249	311257.274	0		173765.657		2										

Mostrando registros del 1 al 10 de un total de 91 registros

Anterior [1](#) [2](#) [3](#) [4](#) [5](#) ... [10](#) Siguiente

© Programa realizado por: Gómez Luna Alejandro

Figura 9: Vista para aplicar métricas de distancia

En este apartado se permite implementar las métricas de distancia euclidiana, de Chebyshev, de Manhattan y de Minkowski para el conjunto de datos proporcionado.

Para el apartado de Ingresar Parámetros se tiene un menú desplegable para seleccionar la métrica que se desee. En caso de seleccionar Minkowski se habilita un apartado para ingresar el parámetro de Lambda, en cualquier otro caso no se muestra dicho apartado. Posteriormente se muestra si se desea calcular toda la matriz de distancias, solamente para un subconjunto de distancias o para una pareja de valores. En caso de seleccionar estas dos últimas opciones se habilita un apartado para seleccionar el rango de valores separado por un guion o la pareja de valores separado por una coma, respectivamente.

Finalmente se muestra la matriz de distancias, que podrá corresponder a un subconjunto o a todos los datos proporcionados. Si se seleccionó una pareja de valores solamente se muestra el resultado del cálculo de la distancia según la métrica que se haya seleccionado.

#### 4. Desglose del apartado MATRIZ DE CORRELACIONES

Correlaciones

PROYECTO IA

VISUALIZAR DATOS

USUARIO FINAL

ALGORITMOS

TABLA CON DATOS

Mostrar 10 registros

Buscar:

VARIABLES	ingresos	gastos_comunes	pago_coche	gastos_otros	ahorros	vivienda	estado_civil	hijos
ingresos	1	0.5602109789	-0.1097803762	-0.1241046977	0.7128893738	0.6147206603	-0.0425555929	-0.024483
gastos_comunes	0.5602109789	1	-0.0543996983	-0.0998806589	0.2094138178	0.204780526	-0.0571524693	-0.072321
pago_coche	-0.1097803762	-0.0543996983	1	0.0106024414	-0.1932994236	-0.0946307533	0.0522390017	-0.044858
gastos_otros	-0.1241046977	-0.0998806589	0.0106024414	1	-0.0643843099	-0.0545766641	-0.0202255894	0.124844
ahorros	0.7128893738	0.2094138178	-0.1932994236	-0.0643843099	1	0.6058356002	-0.0630390192	0.001445
vivienda	0.6147206603	0.204780526	-0.0946307533	-0.0545766641	0.6058356002	1	-0.113420194	-0.141924
estado_civil	-0.0425555929	-0.0571524693	0.0522390017	-0.0202255894	-0.0630390192	-0.113420194	1	0.507608
hijos	-0.0244831085	-0.072321268	-0.0448581641	0.1248447534	0.0014453614	-0.1419243028	0.5076087233	1
trabajo	-0.0388524292	-0.0790954149	0.0189455788	0.0473131866	-0.0238290618	-0.2117895356	0.5895119943	0.699916

Mostrando registros del 1 al 10 de un total de 10 registros

Anterior 1 Siguiente

Figura 10: Primera parte de la vista para el análisis correlacional de variables

En este apartado se podrá realizar el análisis previo de los datos para observar el grado de correlación entre las variables y así, en caso de considerarse necesario, para posteriores algoritmos donde se permita seleccionar variables realizar una reducción de la dimensionalidad, con lo cual disminuya la complejidad computacional y la probabilidad de inconsistencias debido a redundancias entre las variables.

Primero se muestra la matriz de correlaciones, en donde se tiene una tabla con las características similares previamente dichas, donde se puede ordenar cada columna de mayor a menor para observar el grado de correlación de las demás variables respecto a esa columna.

Finalmente, se muestra un mapa de calor con el cual se hace más explícito el grado de correlación entre las variables para poder facilitar el proceso de reducción de la dimensionalidad.

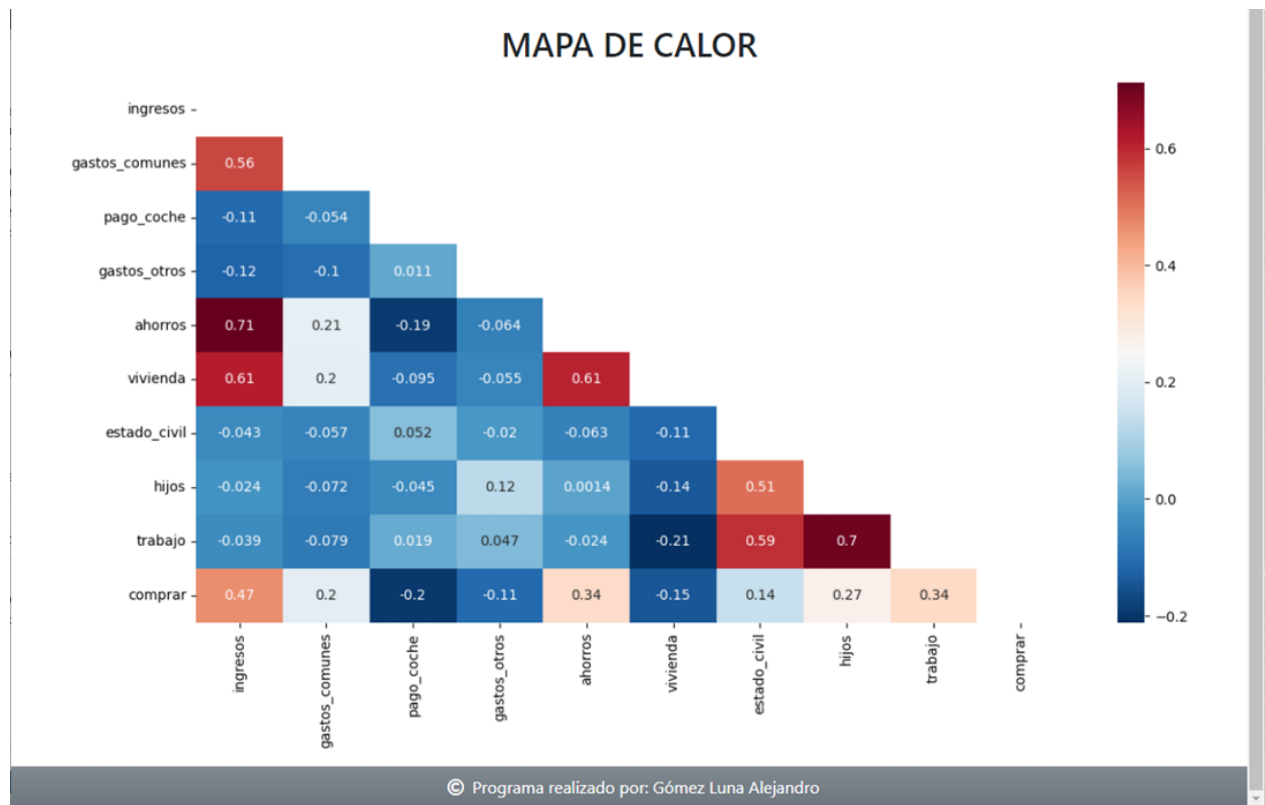


Figura 11: Segunda parte de la vista para el análisis correlacional de variables

## 5. Desglose del apartado CLÚSTERING JERÁRQUICO

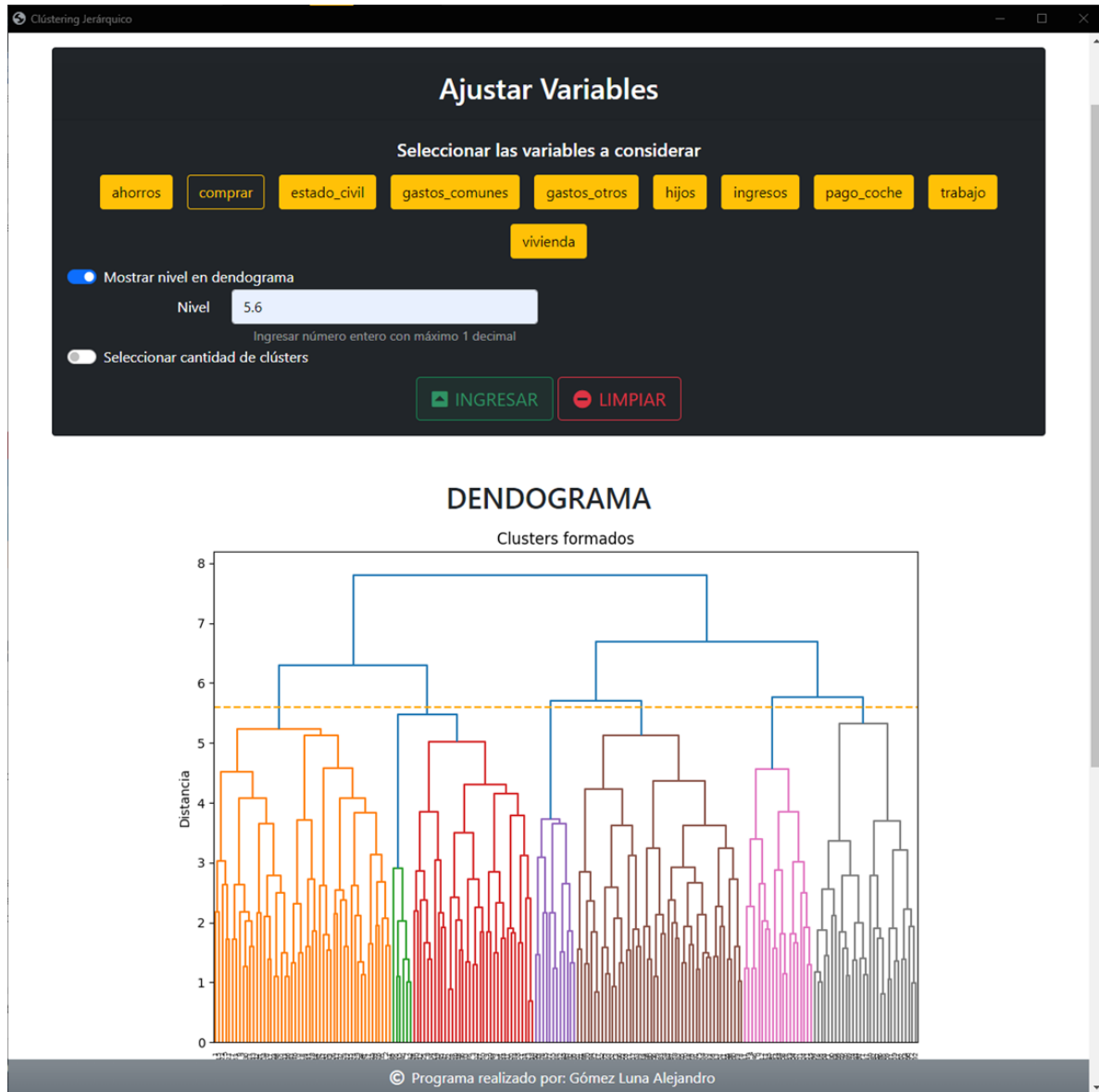


Figura 12: Primera parte de la vista para el algoritmo de clústering ascendente jerárquico

En este apartado se implementa el algoritmo de clústering ascendente jerárquico. Primero, se le solicita al usuario que seleccione las variables a considerar, siendo consideradas todas por defecto. Todas aquellas que no se deseen considerar deberán de deshabilitarse al oprimir el botón con el nombre de la variable que desea descartarse. Posteriormente se ofrecen las opciones de mostrar nivel en dendograma o seleccionar cantidad de clústers, las cuales son mutuamente exclusivas, pues se debe decidir si generar un dendograma para hacer un análisis previo con el cual determinar la cantidad de clústers o ya generar los clústers. Por defecto, si ninguna opción es seleccionada se genera un dendograma, y si se escoge la primera opción entonces se genera el dendograma con una línea pintada para tener una ayuda visual del número de clústers que se forman en cierto nivel.

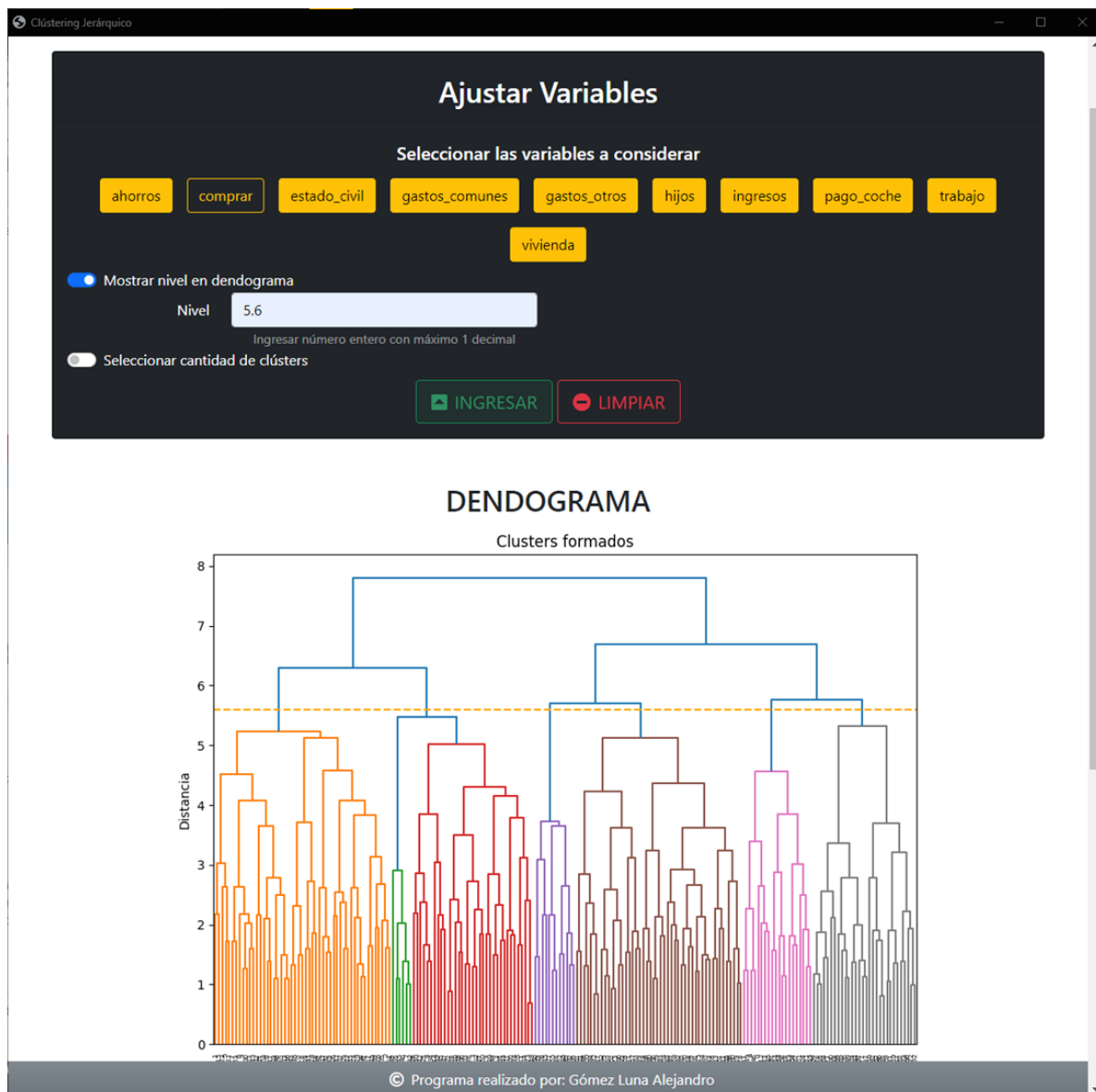


Figura 13: Segunda parte de la vista para el algoritmo de clústering ascendente jerárquico

Una vez que se ha ingresado el número de clústers entonces se procede a mostrar la cantidad de clústers formado con un gráfico de barras, en el cual se visualiza cuántos elementos existen por cada uno de los clústers formados y, también se muestran cada uno de los clústers formados a través de una lista desplegable, donde al abrir una pestaña se muestra el centroide de dicho clúster para cada una de las variables consideradas, junto a un área de texto para generar la respectiva conclusión sobre dicho clúster.

Hasta la parte inferior se presenta un botón para guardar dichas conclusiones. Al hacer click sobre dicho botón se desplegará el explorador de archivos para que nosotros almacenemos las conclusiones generadas en algún directorio de nuestra computadora.



## 6. Desglose del apartado CLÚSTERING PARTICIONAL



Figura 14: Primera parte de la vista para el algoritmo de clústering particional k-means

En este apartado se implementa el algoritmo de clústering particional k-means.

Primero se le solicita al usuario que seleccione las variables a considerar, siendo consideradas todas por defecto. Todas aquellas que no se deseen considerar deberán de deshabilitarse al oprimir el botón con el nombre de la variable que desea descartarse. Posteriormente se ofrecen las opciones de ajustar rango máximo para el método del codo o seleccionar cantidad de clústers, las cuales son mutuamente exclusivas, pues se debe decidir si generar la gráfica de los SSE para diferentes valores de k y así aplicar el método del codo como un análisis previo con el cual determinar la cantidad de clústers o ya generar los clústers.

Por defecto, si ninguna opción es seleccionada se genera la gráfica de los SSE para valores de  $k$  comprendidos entre 2 y 12. En caso de especificarse un valor máximo entonces la gráfica se generará hasta dicho valor.

Debido a la naturaleza de este método, también se ofrece una recomendación de dónde se encuentra el codo según un algoritmo localizador de codo, con lo cual se le ofrece una ayuda al usuario para determinar la cantidad de clústers a formar.



Figura 15: Segunda parte de la vista para el algoritmo de clústering particional k-means

Una vez que se ha ingresado el número de clústers entonces se procede a mostrar la cantidad de clústers formado con un gráfico de barras, en el cual se visualiza cuántos elementos existen por cada uno de los clústers formados y, también se muestran cada uno de los clústers formados a través de una lista desplegable, donde al abrir una pestaña se muestra el centroide de dicho clúster para cada una de las variables consideradas, junto a un área de texto para generar la respectiva conclusión sobre dicho clúster.

Hasta la parte inferior se presenta un botón para guardar dichas conclusiones. Al hacer click sobre dicho botón se desplegará el explorador de archivos para que nosotros almacenemos las conclusiones generadas en algún directorio de nuestra computadora.

## 7. Desglose del apartado REGRESIÓN LOGÍSTICA

Regresión Logística

PROYECTO IA

VISUALIZAR DATOS

USUARIO FINAL

ALGORITMOS ▾

### Ajustar Variables

Seleccionar las variables a considerar

Area Compactness ConcavePoints Concavity Diagnosis FractalDimension IDNumber Perimeter Radius

Smoothness Symmetry Texture

INGRESAR LIMPIAR

### Ajustar Modelo

Seleccionar la variable a predecir

Diagnosis

Seleccionar el porcentaje de datos para prueba

20%

INGRESAR LIMPIAR

Figura 16: Primera parte de la vista para el algoritmo de regresión logística

En este apartado se implementa el método de regresión logística para poder realizar predicciones respecto a una variable clase con respecto a otras variables predictoras.

Primero se le solicita al usuario que seleccione las variables a considerar, siendo consideradas todas por defecto. Todas aquellas que no se deseen considerar deberán de deshabilitarse al oprimir el botón con el nombre de la variable que desea descartarse.

Posteriormente se habilita el bloque para ajustar el modelo, en donde nosotros seleccionamos la variable clase, es decir, sobre la cual queremos realizar predicciones. También podemos seleccionar el porcentaje de datos a utilizar para la fase de pruebas, variando desde un rango del 5 % al 30 %, teniendo saltos de 1 % en 1 %.

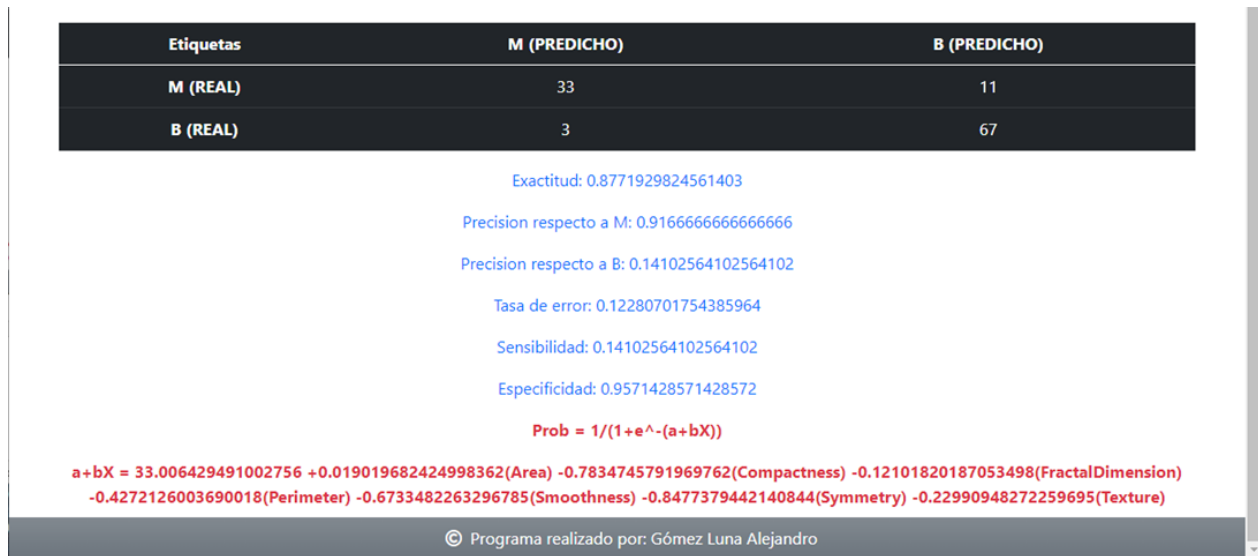


Figura 17: Segunda parte de la vista para el algoritmo de regresión logística

Finalmente se muestra la matriz de confusión con sus respectivos parámetros, así como el modelo de regresión logística que se generó.

Una vez que se ha completado todo este proceso entonces podremos ocupar la pestaña de usuario final.

## 8. Desglose del apartado ÁRBOL DE DECISIÓN

The image shows a web application interface for a Decision Tree algorithm, divided into two main sections: 'Ajustar Variables' and 'Ajustar Modelo'.

**Ajustar Variables:** This section allows users to select variables to consider. It features a grid of buttons for variables: Area, Compactness, ConcavePoints, Concavity, Diagnosis, FractalDimension, IDNumber, Perimeter, Radius, Smoothness, Symmetry, and Texture. Below the grid are two buttons: 'INGRESAR' (green) and 'LIMPIAR' (red).

**Ajustar Modelo:** This section allows users to adjust model parameters. It includes a dropdown menu to select the variable to predict (currently 'Diagnosis'). Below this is a slider to select the percentage of data for testing, currently set at 15%. There are two blue informational buttons: 'En caso de no querer ajustar los parámetros para la construcción del árbol no ingresar ningún valor' and 'Por defecto se generará un árbol de decisión para regresión'. Below these are three input fields for 'Max-Depth', 'Min-samples-split', and 'Min-samples-leaf', each with a default value of '00' and a descriptive tooltip. At the bottom left, there is a toggle switch for 'Árbol para clasificación' (currently on). At the bottom right are 'INGRESAR' (green) and 'LIMPIAR' (red) buttons. The footer indicates '© Programa realizado por: Gómez Luna Alejandro'.

Figura 18: Primera parte de la vista para el algoritmo de árboles de decisión

En este apartado se implementa el método de regresión logística para poder realizar predicciones respecto a una variable clase con respecto a otras variables predictoras.

Primero se le solicita al usuario que seleccione las variables a considerar, siendo consideradas todas por defecto. Todas aquellas que no se deseen considerar deberán de deshabilitarse al oprimir el botón con el nombre de la variable que desea descartarse.

Posteriormente se habilita el bloque para ajustar el modelo, en donde nosotros seleccionamos la variable clase, es decir, sobre la cual queremos realizar predicciones. También podemos seleccionar el porcentaje de datos a utilizar para la fase de pruebas, variando desde un rango del 5 % al 30 %, teniendo saltos de 1 % en 1 %.

En este mismo bloque se muestran dos avisos para el ajuste de los parámetros del árbol de decisión, en donde podemos ajustar la profundidad máxima del árbol, la mínima cantidad de elementos que debe de haber en los nodos de decisión y la mínima cantidad de elementos que debe de haber en los nodos hoja. Por defecto no se tiene una profundidad máxima, la cantidad de elementos mínima en los nodos de decisión es de 2 y la cantidad de elementos mínima en los nodos hoja es de 1. Un último ajuste es decidir si queremos generar un árbol de decisión para clasificación. En caso de no habilitar la opción, entonces se generará un árbol de decisión para regresión.

#	Variable	Importancia
0	Area	0.6376190399
1	Compactness	0.0842415445
7	Texture	0.0786857777
5	Smoothness	0.0751500593
3	Perimeter	0.0503928489
6	Symmetry	0.0460698687
2	FractalDimension	0.0234603059
4	Radius	0.0043805551

Etiquetas	M	B
M	23	4
B	4	55

Exactitud: 0.9069767441860465

Precision respecto a M: 0.8518518518518519

Precision respecto a B: 0.06779661016949153

Tasa de error: 0.09302325581395349

Sensibilidad: 0.06779661016949153

Especificidad: 0.9322033898305084

Figura 19: Segunda parte de la vista para el algoritmo de árboles de decisión

Para la opción de árbol de decisión para clasificación se muestra la matriz de confusión con sus respectivos parámetros.

#	Variable	Importancia
3	Radius	0.9990696193
0	Compactness	0.000328869
2	Perimeter	0.0002579978
4	Smoothness	0.0001302417
5	Symmetry	0.0001060756
6	Texture	0.0000807531
1	FractalDimension	0.0000264434

MAE: 11.973255813953497

MSE: 984.1508139534884

RMSE: 31.371178077233385

Score: 0.9919684139560082

Figura 20: Tercera parte de la vista para el algoritmo de árboles de decisión

Para la opción de árbol de decisión para regresión únicamente se muestran los parámetros de MAE, MSE, RMSE y el Score del modelo.

En ambos casos se muestra una tabla en donde se pueda visualizar a cada una de las variables que se involucraron para la generación del árbol, así como su porcentaje de importancia para realizar el pronóstico.

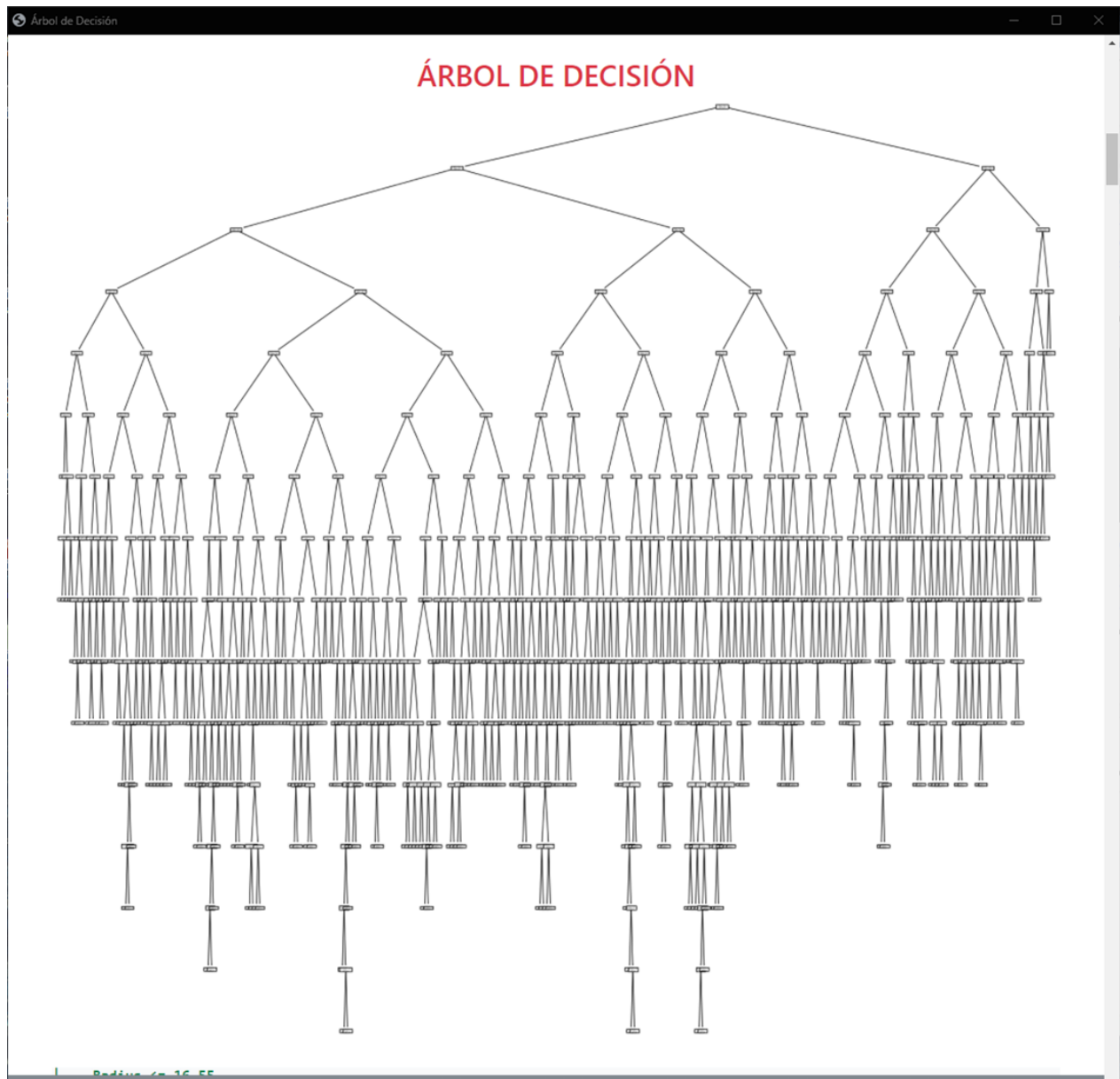


Figura 21: Cuarta parte de la vista para el algoritmo de árboles de decisión

Posteriormente se genera el árbol de decisión de manera general, para visualizar el nivel de profundidad que tiene para así tener mayores indicios de si existe o no un sobreajuste.



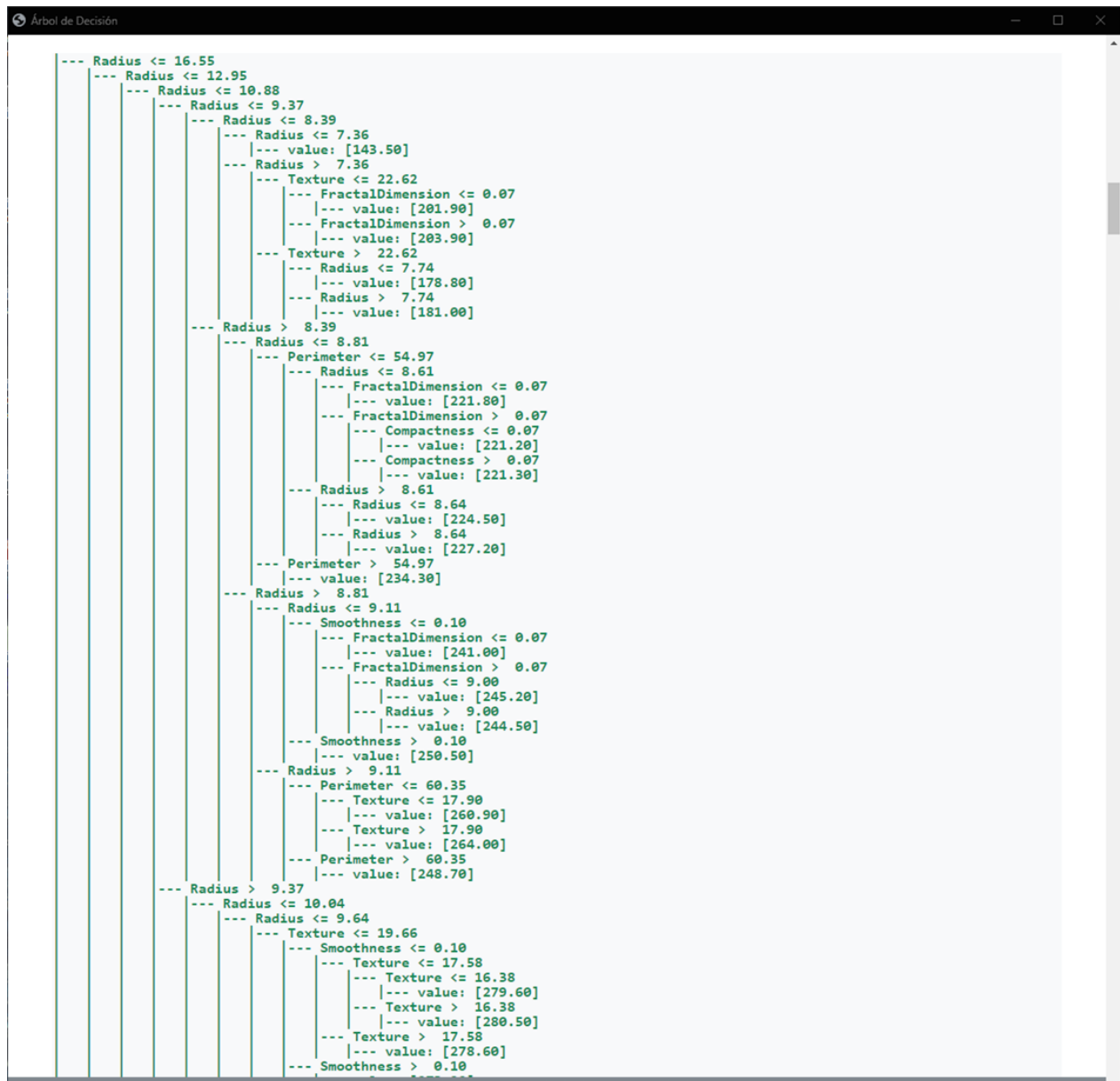


Figura 22: Quinta parte de la vista para el algoritmo de árboles de decisión

Finalmente se muestra el árbol de decisión generado con formato de texto, con lo cual es más fácil visualizar los nodos que existen en conjunto con sus respectivos criterios que se realizaron para conformarlos.

## 9. Desglose del apartado *USUARIO FINAL*

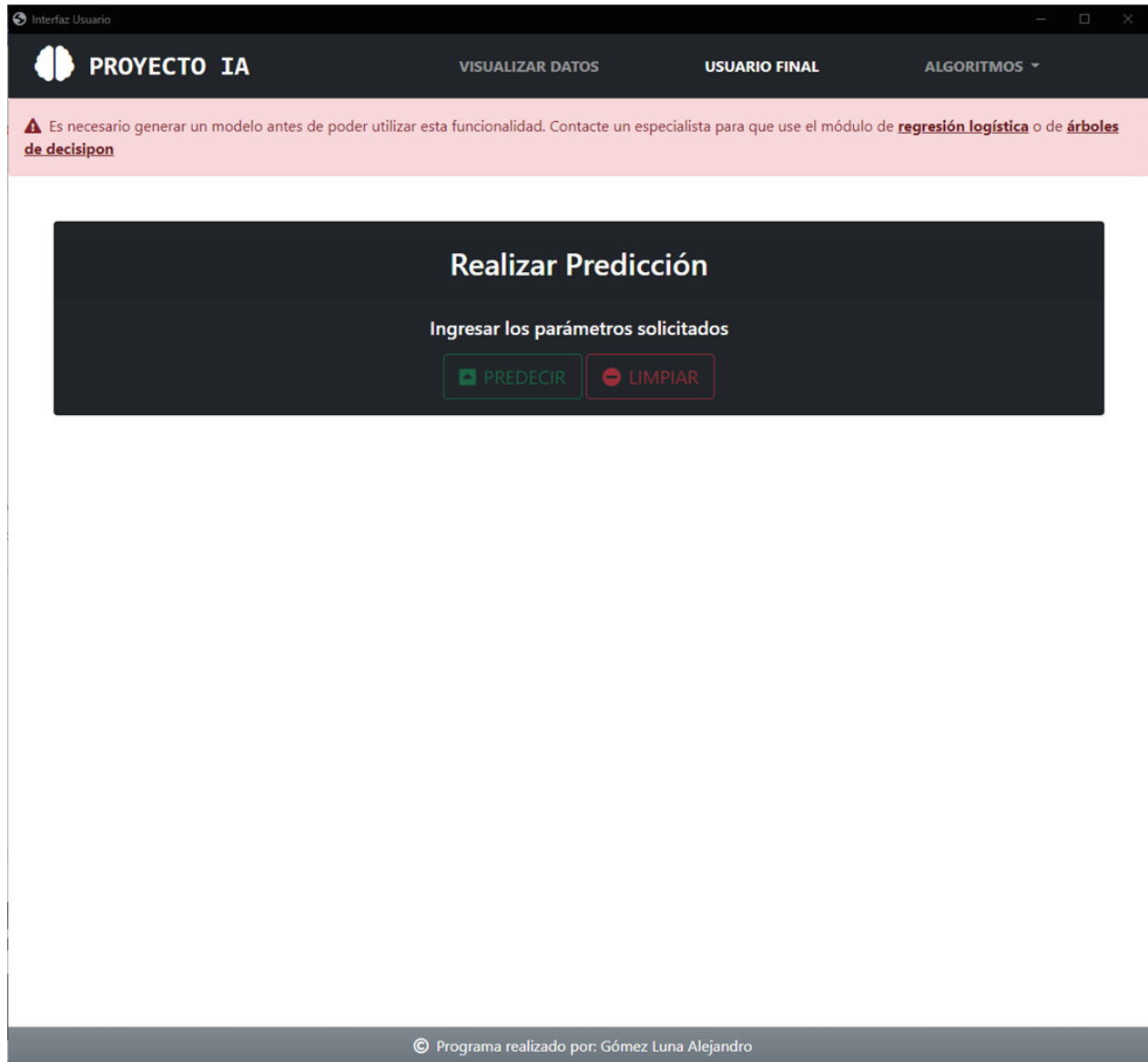


Figura 23: Vista para el usuario final sin el modelo asociado

Esta última vista podrá visualizarse de dos maneras distintas.

La primera forma es como la mostrada en la imagen, en donde se muestra una leyenda de advertencia, indicándole al usuario que quiera utilizar esta funcionalidad que primero debe de generar algún modelo mediante los algoritmos de regresión logística o de árboles de decisión para poder ingresar parámetros y realizar predicciones, ya que si no se cuenta con ningún modelo entonces no se pueden realizar ninguna clase de pronóstico. Se ofrecen los enlaces para que se acceda directamente a aplicar estos algoritmos.

Interfaz Usuario

PROYECTO IA VISUALIZAR DATOS USUARIO FINAL ALGORITMOS

### Realizar Predicción

Ingresar los parámetros solicitados

Area	1	Compactness	1
	Ingresar unicamente números decimales		Ingresar unicamente números decimales
ConcavePoints	1	FractalDimension	1
	Ingresar unicamente números decimales		Ingresar unicamente números decimales
Perimeter	1	Smoothness	1
	Ingresar unicamente números decimales		Ingresar unicamente números decimales
Symmetry	1	Texture	1
	Ingresar unicamente números decimales		Ingresar unicamente números decimales

PREDECIR LIMPIAR

El valor que se predijo fue: B

La predicción tiene una exactitud de: 0.9298245614035088

© Programa realizado por: Gómez Luna Alejandro

Figura 24: Vista para el usuario final con el modelo asociado

Esta es la segunda forma de la vista, en donde ya se ha generado un modelo a través de alguno de los dos algoritmos anteriores, por lo tanto se pueden observar en el recuadro los distintos parámetros a ingresar, los cuales se toman como base a las variables que se seleccionaron para crear al modelo. Una vez que se han ingresado todos estos valores en el recuadro entonces se muestra a través de un mensaje de texto el valor que se predijo, así como la exactitud con la cual se hizo la predicción.

## Proceso de ejecución

- I. Descargar la aplicación llamada interface y guardarla en algún directorio de la computadora.
- II. Una vez descargada se da doble click para ejecutarla. Es importante recordar que todos los archivos que vayamos a utilizar en los distintos algoritmos deben de estar en el mismo directorio donde se ha descargado la aplicación.

## Requisitos de ejecución

- Contar con un explorador web, siendo Google Chrome el recomendado.
- Utilizar Windows 7/8/10/11 como sistema operativo.

*\*Nota: Al momento de realizar la descarga de la aplicación es bastante probable que aparezca es un tipo de malware, esto debido a que Python Eel hace uso de un Web Socket para poder iniciar la aplicación y que todo se ejecute como si se estuviera ejecutado en un servidor local en la computadora, por lo que este tipo de comportamiento sigue un patrón común de muchos tipos de malware y por ende es catalogado como si se tratara de una aplicación peligrosa.*

## Conclusiones

Para este proyecto se logró el objetivo de desarrollar una interfaz gráfica con la cual se le pueda facilitar a algún especialista en el campo de la IA el proceso de aplicación de distintos algoritmos con los cuales se puedan analizar conjuntos de datos con la finalidad de obtener información de interés que será utilizada para distintos fines.

Con el desarrollo de esta interfaz se logró reforzar aún más los conceptos teóricos vistos en clase, así como tener un mayor entendimiento de los algoritmos que se estuvieron aplicando a lo largo de las distintas prácticas que se realizaron en el semestre, pues para poder implementar los algoritmos en esta interfaz se necesitó analizar la forma en cómo trabajan, en donde se necesitaba conocer el tipo de dato que necesitaban como entrada, la forma en cómo devolvían los resultados para darles un correcto formato y mostrarlos en pantalla, entre otras cuestiones.

Además de este aprendizaje teórico, se logró profundizar un poco más en lenguajes de programación como JavaScript y Python, con los cuales se logró la mayoría del desarrollo de la interfaz. El poder comprender las bondades que ofrece cada lenguaje fue de suma importancia para lograr un proceso de desarrollo más fluido y sencillo.

Si bien el proceso de desarrollo en muchas ocasiones llegó a ser algo tardado o bastante difícil, me gustaría resaltar que el hecho de haber creado un esquema de trabajo y apegarse en la mayoría de las ocasiones a este, permitió tener la organización necesaria para poder desarrollar esta interfaz con el tiempo necesario para ir cumpliendo con cada uno de los requerimientos que se iban solicitando semana tras semana.

Es así como el desarrollo de esta interfaz proporcionó un aprendizaje teórico y práctico respecto a los algoritmos de reglas de asociación, clústering jerárquico ascendente y particional k-means, regresión logística, árboles de decisión, si no que también se entendió mejor la importancia que tiene el análisis previo de los datos para conocer su naturaleza y características con la finalidad de aplicar los algoritmos de mejor manera. También el desarrollo de esta interfaz permitió ahondar más en otros lenguajes de programación y en bibliotecas desarrolladas en estos, las cuales ofrecieron las herramientas necesarias para crear una interfaz más eficiente y funcional.