

Proyecto Final - Interfaz Gráfica de Usuario (GUI)

Jiménez Juárez Jesús

Universidad Nacional Autónoma de México (UNAM), Facultad de Ingeniería

Minería de datos - Semestre 2021-2

jesus.jimenez@ingenieria.unam.edu

10 de agosto de 2021

1. Introducción

Ha habido un gran cambio en el mundo con lo que se ha creado y descubierto a lo largo de los años. En la actualidad, se tienen desarrolladas tecnologías que usan Inteligencia Artificial que son útiles para cualquier sector, ya sea en el área de salud, económico, industrial, político, financiero, entre otros.

Una de las ramas que sin duda alguna ha sido relevante en el campo de la Inteligencia Artificial es la minería de datos, la cual se utiliza para poder analizar grandes cantidades de datos con el fin de extraer información que sea de utilidad para las investigaciones que se estén realizando.

Este proyecto no es la excepción con lo antes mencionado, esto porque se va a realizar una Interfaz Gráfica de Usuario (GUI) con algunos algoritmos que se vieron en clase, que será útil para aquellas personas que estén elaborando la exploración y análisis de datos de un tema en específico a cierta población.

Dichos algoritmos, que se van a explicar más adelante, son:

- Análisis Exploratorio de datos.
- Selección de características, enfocado en Componentes principales.
- Cluster Particional.
- Regresión Logística.

2. Objetivo

Implementar una interfaz gráfica de usuario (GUI) con ciertos algoritmos en específico vistos en clases con las cuales se busca analizar y explorar los datos con diferentes fines, ya sea para reducir la cantidad de variables de un conjunto de datos, organizar los datos en grupos o para realizar pronósticos.

3. Marco Teórico

- **Análisis Exploratorio de datos:** Es un proceso de investigación en el que se usan estadísticas de resumen y herramientas gráficas para llegar a conocer los datos y comprender lo que se puede averiguar de ellos. Las actividades que realiza son las siguientes:

- Paso 1: Descripción de la estructura de los datos.
- Paso 2: Identificación de datos faltantes.
- Paso 3: Detección de valores atípicos.
- Paso 4: Identificación de relaciones entre pares variables.

- **Componentes Principales:** El análisis de componentes principales es un algoritmo matemático para disminuir la cantidad de variables de conjuntos de datos, mientras se conserva un gran número de información posible. Estos componentes principales (vectores o factores) son combinaciones lineales, no correlacionadas entre sí, que guardan la mayor cantidad de varianza.

- **Cluster:** Consiste en la segmentación y delimitación de grupos de elementos, generalmente representados en forma de datos, que son unidos por características en común que estos comparten (aprendizaje no supervisado). Para hacer clustering, se debe saber el grado de similitud de los elementos. El algoritmo particional, también llamado como de particiones, organiza los elementos dentro de k clústeres o grupos.

- **Regresión logística:** Es un algoritmo de aprendizaje supervisado cuyo objetivo es predecir valores binarios (0 o 1).

- Se usa la misma estructura que la regresión lineal, pero se transforma la variable respuesta (0 o 1) en una probabilidad.
- Para esta transformación se utiliza la función logística (conocida también como sigmoide).

4. Implementación de la Interfaz Gráfica de Usuario (GUI)

En un principio, se pensó realizar la Interfaz Gráfica de Usuario con HTML, CSS y Flask, este último es un microframework de Python para crear aplicaciones web. En lo personal, en el campo del desarrollo web no soy un experto del todo y me costó trabajo implementar los algoritmos dentro de las herramientas antes mencionadas. Es por eso que pensé mejor en utilizar en un lenguaje de programación que conociera mucho más a detalle, el cual es Python.

Dentro de él, existen una gran cantidad de librerías que proporcionan diferentes aplicaciones. Una de las más importantes, la cual es Tkinter, ayuda a realizar interfaces de usuario fáciles de programar y anteriormente ya se había trabajado con ello para distintos proyectos que se habían realizado.

Otras de las librerías que se utilizaron y que fueron de gran aportación para el desarrollo de este proyecto son los siguientes:

- **pandas:** es una librería de Python especializada en el manejo y análisis de estructuras de datos. Las principales características son:
 - Permitir leer y escribir de manera sencilla ficheros en formato CSV, Excel y bases de datos SQL.
 - Permitir acceder a los datos por medio de índices o nombres para filas y columnas.
 - Ofrecer métodos para reordenar, dividir y combinar conjuntos de datos.
- **numpy:** NumPy es una librería de Python especializada en el cálculo numérico y el análisis de datos, especialmente para un gran volumen de datos.

Incorpora una nueva clase de objetos llamados arrays que permite representar colecciones de datos de un mismo tipo en varias dimensiones, y funciones muy eficientes para su manipulación.

- **sklearn:** es una librería de Python que integra algoritmos clásicos de aprendizaje automático.

Cuenta con algoritmos de clasificación, regresión, clustering y reducción de dimensionalidad. Además, presenta la compatibilidad con otras librerías de Python como NumPy y matplotlib.

La gran variedad de algoritmos y utilidades de esta librería la convierten en la herramienta básica para empezar a programar y estructurar los sistemas de análisis de datos y modelado estadístico.

- **matplotlib.pyplot**: es una colección de funciones que hacen que matplotlib funcione como MATLAB. Cada función de pyplot realiza algún cambio en una figura; por ejemplo, crear una figura, crear un área de trazado en una figura, trazar algunas líneas en un área de trazado, entre otras más.

- **seaborn**: es una librería de visualización de datos construida sobre matplotlib y estrechamente integrada con las estructuras de datos de pandas en Python. La visualización es la parte central de Seaborn que ayuda en la exploración y comprensión de los datos.

5. Funcionamiento

Lo que hay que tomar en cuenta desde un principio es que Python debe estar instalado en la computadora o laptop para poder ejecutar el programa. Se debe verificar con el comando `python --version` en la terminal del sistema operativo que se esté manejando. En caso de no tenerlo, se recomienda la siguiente liga para descargarlo (<https://www.python.org/downloads/>). Marque todas las casillas al momento de ir configurando la instalación, como se muestra en las siguientes imágenes.



Imagen 1. Configuración al instalar Python (I).

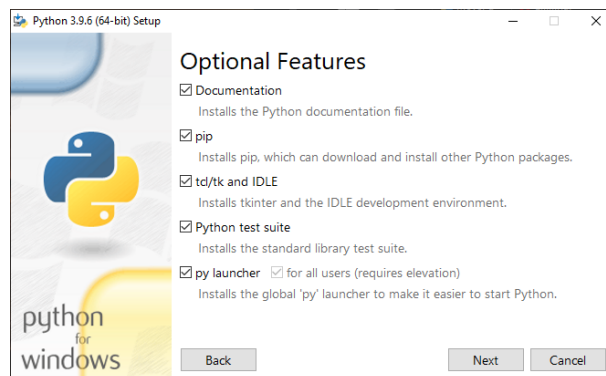
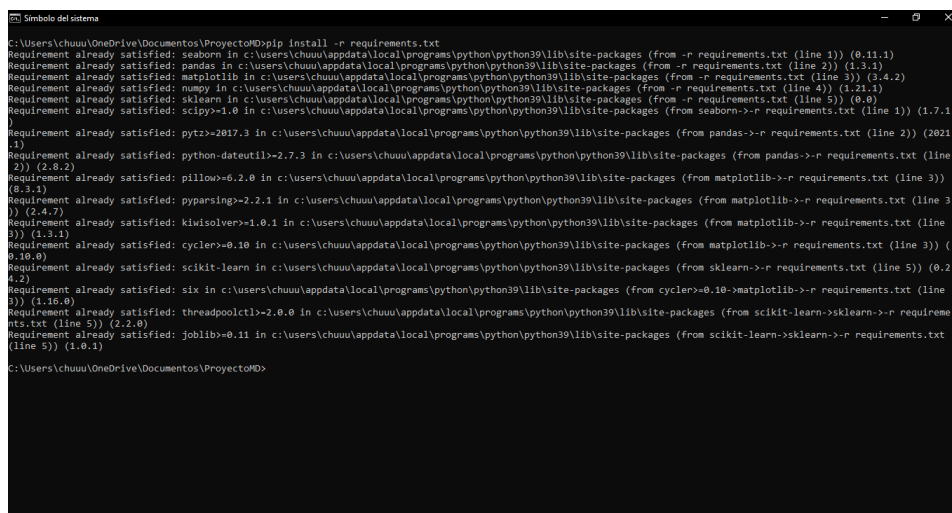


Imagen 2. Configuración al instalar Python (II).

Ya hecho lo anterior, se debe ingresar al directorio donde se tienen guardado los archivos InterfazMD.py y requirements.txt. Se requieren de varias librerías que posiblemente no se tienen instaladas, es por ello que ejecutamos el siguiente comando:

pip install -r requirements.txt



```
Simbolo del sistema
C:\Users\chuuu\OneDrive\Documents\ProyectoMD>pip install -r requirements.txt
Requirement already satisfied: seaborn in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from -r requirements.txt (line 1)) (0.11.1)
Requirement already satisfied: pandas in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from -r requirements.txt (line 2)) (1.3.1)
Requirement already satisfied: matplotlib in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from -r requirements.txt (line 3)) (3.4.2)
Requirement already satisfied: numpy in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from -r requirements.txt (line 4)) (1.21.1)
Requirement already satisfied: sklearn in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from -r requirements.txt (line 5)) (0.0)
Requirement already satisfied: scipy>=1.0 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from seaborn->-r requirements.txt (line 1)) (1.7.1)
Requirement already satisfied: pytz>=2017.3 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from pandas->-r requirements.txt (line 2)) (2021.1)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from pandas->-r requirements.txt (line 2)) (2.8.2)
Requirement already satisfied: pillow>=6.2.0 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (8.3.1)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (2.4.7)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (1.3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from matplotlib->-r requirements.txt (line 3)) (0.10.0)
Requirement already satisfied: scikit-learn in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from sklearn->-r requirements.txt (line 5)) (0.24.2)
Requirement already satisfied: six in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from cycler>=0.10->matplotlib->-r requirements.txt (line 3)) (1.16.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from scikit-learn->sklearn->-r requirements.txt (line 5)) (2.2.0)
Requirement already satisfied: joblib>=0.11 in c:\users\chuuu\appdata\local\programs\python\python39\lib\site-packages (from scikit-learn->sklearn->-r requirements.txt (line 5)) (1.0.1)
C:\Users\chuuu\OneDrive\Documents\ProyectoMD>
```

Imagen 3. Ejecución del comando pip install -r requirements.txt.

Ahora lo que se debe de escribir en la línea de comando es lo siguiente para poder ejecutar el proyecto.
python InterfazMD.py



Imagen 4. Ejecución del comando python InterfazMD.py.

Aparecerá lo que se muestra en la imagen anterior, en donde lo que se muestra en esta ventana son las instrucciones que se deben de seguir, los cuales son las siguientes:

1. Subir un archivo, ya sea de extensión csv o txt.
2. Dar click en Avanzar después de haber seleccionado el archivo; en caso de querer salir del programa, dar click en Salir del programa
3. En la siguiente ventana, seleccionar un algoritmo que se encuentra en la barra de menú.
4. Dentro de las interfaces de los algoritmos, seguir los pasos que se indican.

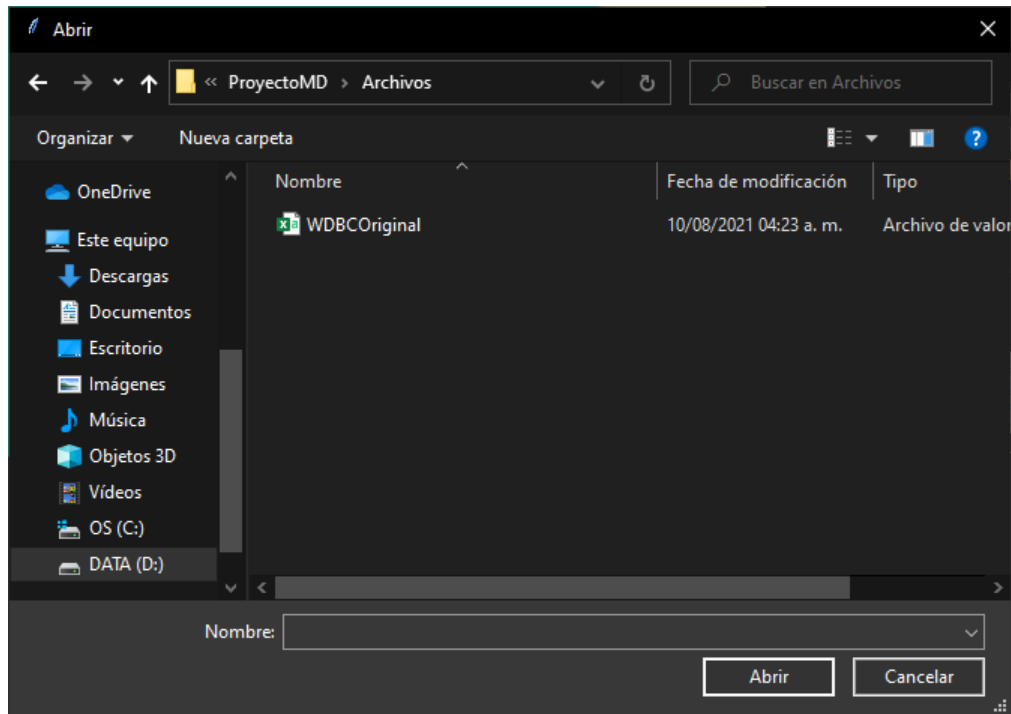


Imagen 5. Ventana externa que aparece al momento de dar en el botón Abrir Archivo.

También hay que mencionar que si el usuario no sube un archivo, se va a mostrar un mensaje en donde se mencione que no se ha subido un archivo. En caso de que se quiera subir un archivo con diferente extensión, el programa no lo dejará y se le dará aviso oportuno de ello para que pueda subir el archivo con la extensión correcta.

Al momento de cargar el archivo y dar click en Avanzar, se deberá desplegar la siguiente pantalla:

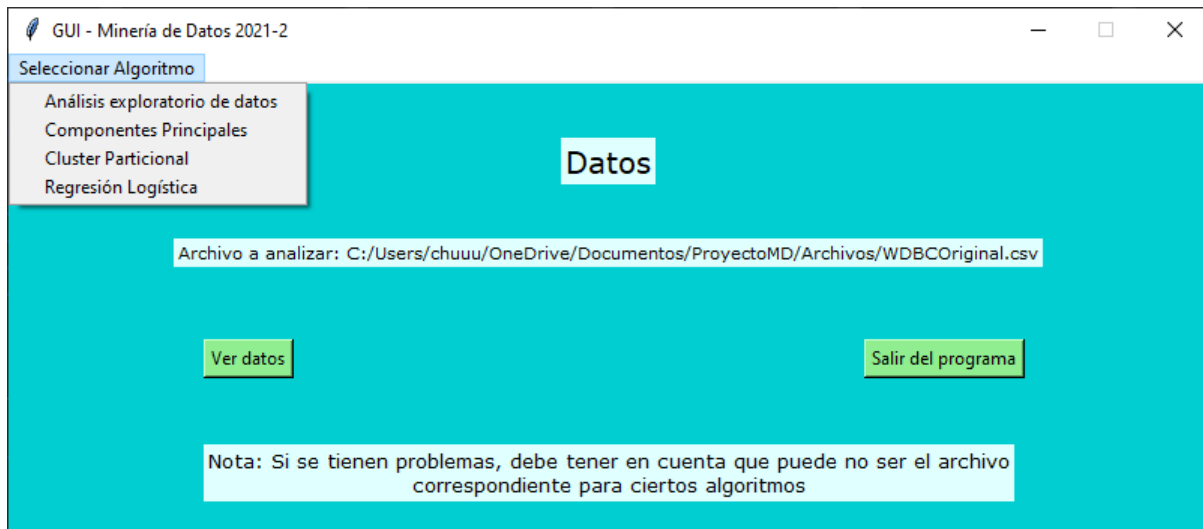


Imagen 6. Ventana donde se muestra el menú de Seleccionar Algoritmo, así como poder visualizar los datos.

Al momento de dar click en Ver Datos, se mostrarán los datos del archivo que se ha cargado.

IDNumber	Diagnosis	Radius	Texture	Perimeter	Area	Smoothness
0	P-842302	M	17.99	10.38	122.80	1001.0
1	P-842517	M	20.57	17.77	132.90	1326.0
2	P-84300903	M	19.69	21.25	130.00	1203.0
3	P-84348301	M	11.42	20.38	77.58	386.1
4	P-84358402	M	20.29	14.34	135.10	1297.0
...
564	P-926424	M	21.56	22.39	142.00	1479.0
565	P-926682	M	20.13	28.25	131.20	1261.0
566	P-926954	M	16.60	28.08	108.30	858.1
567	P-927241	M	20.60	29.33	140.10	1265.0
568	P-92751	B	7.76	24.54	47.92	181.0

Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
0	0.27760	0.30010	0.14710	0.2419
1	0.07864	0.08690	0.07017	0.1812
2	0.15990	0.19740	0.12790	0.2069
3	0.28390	0.24140	0.10520	0.2597
4	0.13280	0.19800	0.10430	0.1809
...
564	0.11590	0.24390	0.13890	0.1726
565	0.10340	0.14400	0.09791	0.1752
566	0.10230	0.09251	0.05302	0.1590
567	0.27700	0.35140	0.15200	0.2397
568	0.04362	0.00000	0.00000	0.1587

[569 rows x 12 columns]

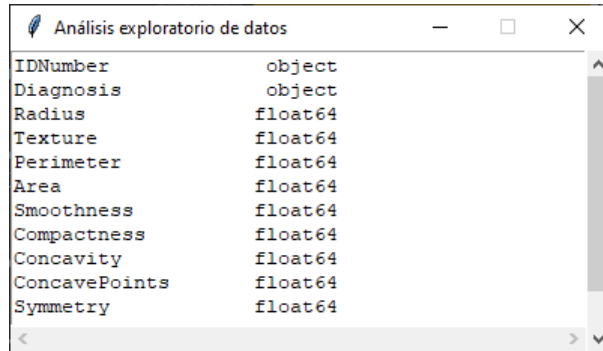
Imagen 7. Datos cargados.

Al elegir el algoritmo Análisis Exploratorio de Datos, se abrirá una nueva ventana el cual mostrará todos los pasos que se mencionaron en el Marco Teórico.



Imagen 8. Ventana del algoritmo Análisis Exploratorio de Datos.

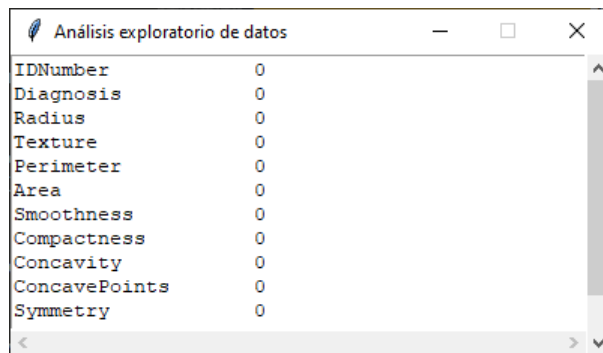
En el paso 1, donde se hace una descripción de la estructura de los datos, se observa el conteo de filas y columnas, así como un botón el cual va a desplegar una ventana con el tipo de variable de los datos.



Variable	Tipo de dato
IDNumber	object
Diagnosis	object
Radius	float64
Texture	float64
Perimeter	float64
Area	float64
Smoothness	float64
Compactness	float64
Concavity	float64
ConcavePoints	float64
Symmetry	float64

Imagen 9. Ventana que se abre al dar click en el botón Ver Tipo de variable de los datos.

En el paso 2, está la identificación de datos faltantes, en donde se observan los variables que tienen valores nulos con el botón Ver variable nulos (que están vacíos).



Variable	Valores nulos
IDNumber	0
Diagnosis	0
Radius	0
Texture	0
Perimeter	0
Area	0
Smoothness	0
Compactness	0
Concavity	0
ConcavePoints	0
Symmetry	0

Imagen 10. Ventana que se abre al dar click en el botón Ver variable nulos (que están vacíos).

En el paso 3, la detección de valores atípicos se divide en dos características, la numérica y la nominal. En la primera de ellas, se tienen dos botones, en uno de ellos se abrirá una ventana en donde aparecerá el histograma de los datos que se cargaron.

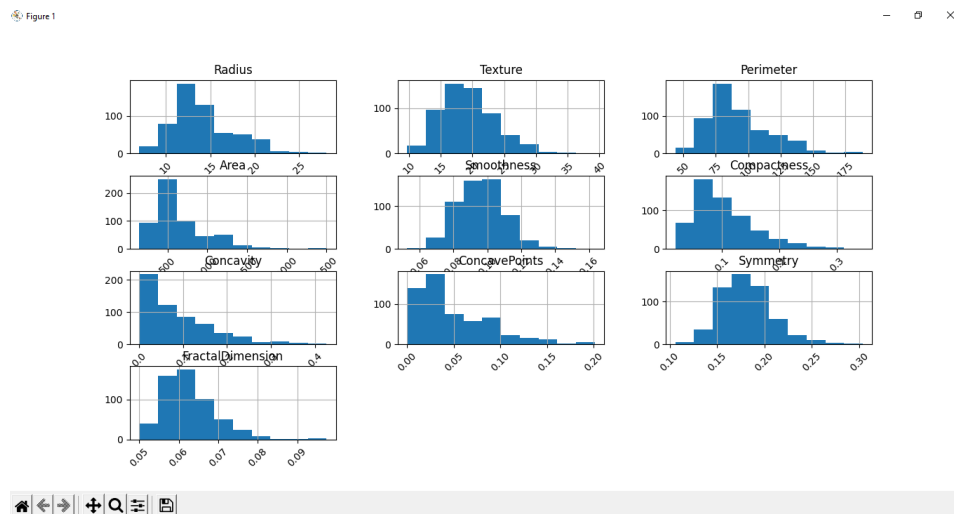


Imagen 11. Histograma de los datos cargados.

Además, en el botón Ver información estadísticos, se va a ver en pantalla la información estadística de cada uno de los datos cargados, es decir, la media, mediana, máximo, mínimo, entre otras más.

	Radius	Texture	Perimeter	Area	Smoothness	\
count	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000
mean	14.127292	19.289649	91.969033	654.889104	0.096360	
std	3.524049	4.301036	24.298981	351.914129	0.014064	
min	6.981000	9.710000	43.790000	143.500000	0.052630	
25%	11.700000	16.170000	75.170000	420.300000	0.086370	
50%	13.370000	18.840000	86.240000	551.100000	0.095870	
75%	15.780000	21.800000	104.100000	782.700000	0.105300	
max	28.110000	39.280000	188.500000	2501.000000	0.163400	

	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
count	569.000000	569.000000	569.000000	569.000000	569.000000
mean	0.104341	0.088799	0.048919	0.181162	0.062798
std	0.052813	0.079720	0.038803	0.027414	0.007060
min	0.019380	0.000000	0.000000	0.106000	0.049960
25%	0.064920	0.029560	0.020310	0.161900	0.057700
50%	0.092630	0.061540	0.033500	0.179200	0.061540
75%	0.130400	0.130700	0.074000	0.195700	0.066120
max	0.345400	0.426800	0.201200	0.304000	0.097440

Imagen 12. Información estadística de los datos.

En la segunda parte de las características, como se mencionó anteriormente, se tiene las nominales, la cual únicamente se tiene la información sobre el conteo de los registros object. Al dar click, aparecerá lo siguiente:

	IDNumber	Diagnosis
count	569	569
unique	569	2
top	P-842302	B
freq	1	357

Imagen 13. Información sobre el conteo de los registros object.

En el último paso, se tiene la identificación de relaciones entre pares variables, esto quiere decir que vamos a hacer uso de la matriz o la gráfica de correlación. Cuando se pulsa en el botón Ver Gráfica de correlación, se mostrará en pantalla lo que se ve en la imagen 14.

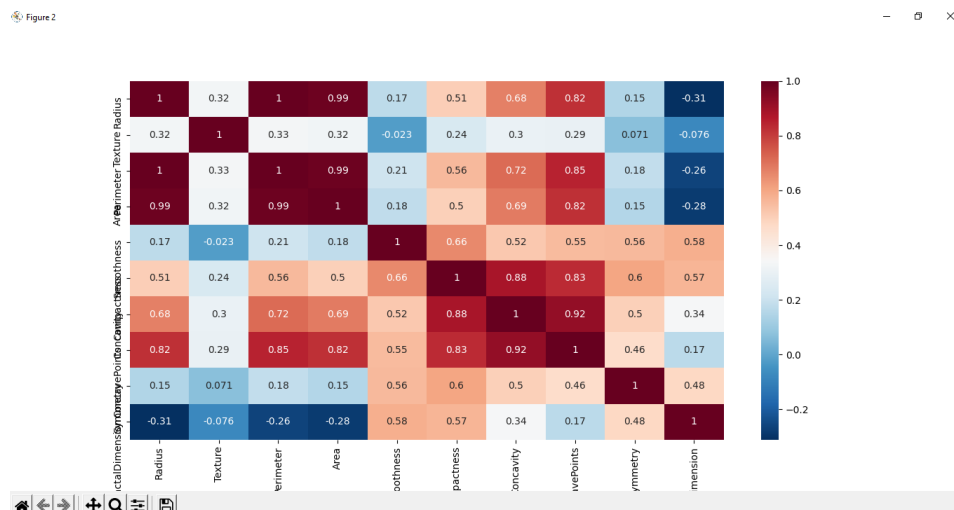


Imagen 14. Matriz de correlación de las variables.

Hay que mencionar que la decisión de cómo es que el usuario desea utilizar los datos quedará a su criterio.

Pasando el segundo algoritmo, el cual es Componentes Principales, aparecerá la siguiente ventana, en la cual se van a mostrar las instrucciones correspondientes que se deben de seguir a detalle, así como las variables del archivo que se está utilizando para su análisis y estudio.

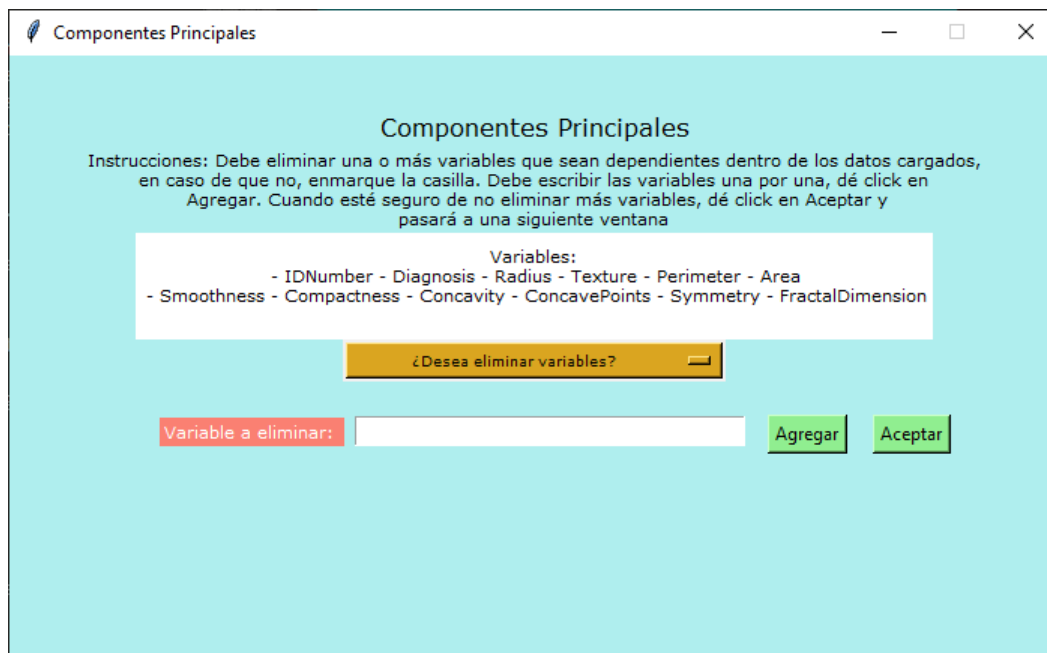


Imagen 15. Ventana del algoritmo Componentes Principales.

Lo que se debe de hacer, es elegir si se va o no eliminar variables dependientes, las cuales son las que se utilizan en la parte del pronóstico. En caso de no eliminarse, puede tener el programa problemas al momento de ejecutar, ya que tienen que ser valores numéricos los que se deben de manejar en este algoritmo. Es por ello que, si se da este error, el programa mandará un mensaje en la cual no se podrá realizar el algoritmo.

Entonces, cuando se tiene que eliminar dichas variables, lo que se debe de hacer es escoger que sí se desean eliminar variables, escribir una a una en la parte blanca, dar click en Agregar y posteriormente en Aceptar.

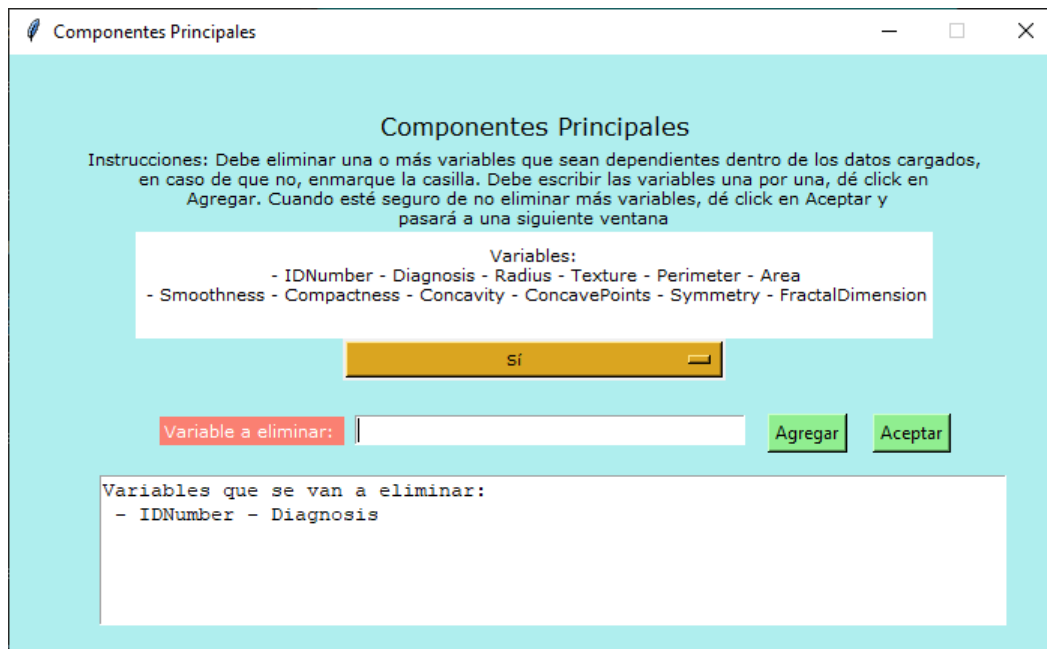


Imagen 16. Ejemplo de cómo se deben eliminar variables.

Si el proceso fue correcto, se va a abrir una nueva ventana en donde se van a mostrar los resultados de los datos que se han cargado.

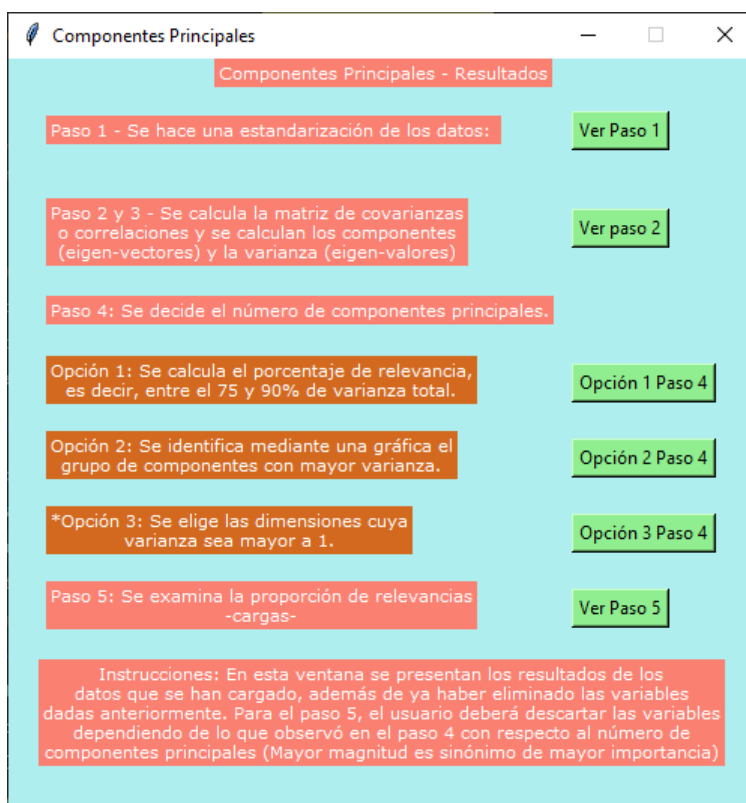


Imagen 17. Ventana de resultados - Componentes Principales.

En el paso 1, se hace una estandarización de los datos, en la cual se abrirá al momento de dar click en Ver Paso 1 los datos normalizados.

	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	ConcavePoints	Symmetry	FractalDimension
0	1.097064	-2.073335	1.269934	0.984375	1.568466	3.283515	2.652874	2.532475	2.217515	2.255747
1	1.829821	-0.353632	1.685955	1.908708	-0.826962	-0.487072	-0.023846	0.548144	0.001392	-0.868652
2	1.579888	0.456187	1.566503	1.558884	0.942210	1.052926	1.363478	2.037231	0.939685	-0.398008
3	-0.768909	0.253732	-0.592687	-0.764464	3.283553	3.402909	1.915897	1.451707	2.867383	4.910919
4	1.750297	-1.151816	1.776573	1.826229	0.280372	0.539340	1.371011	1.428493	-0.009560	-0.562450
...
564	2.110995	0.721473	2.060786	2.343856	1.041842	0.219060	1.947285	2.320965	-0.312589	-0.931027
565	1.704854	2.085134	1.615931	1.723842	0.102458	-0.017833	0.693043	1.263669	-0.217664	-1.058611
566	0.702284	2.045574	0.672676	0.577953	-0.840484	-0.038680	0.046588	0.105777	-0.809117	-0.895587
567	1.838341	2.336457	1.982524	1.735218	1.525767	3.272144	3.296944	2.658866	2.137194	1.043695
568	-1.808401	1.221792	-1.814389	-1.347789	-3.112085	-1.150752	-1.114873	-1.261820	-0.820070	-0.561032

[569 rows x 10 columns]

Imagen 18. Estandarización de los datos.

En el paso 2 y 3, se calcula la matriz de covarianzas o correlaciones y se calculan los componentes (eigen-vectores) y la varianza (eigen-valores).

	0	1	2	3	4	5	6	7	8	9
0	5.224155	3.204428	-2.171340	-0.169276	1.514252	0.113123	-0.344684	-0.231932	-0.021982	-0.011258
1	1.728094	-2.540839	-1.019679	0.547539	0.312330	-0.935634	0.420922	-0.008343	-0.056171	-0.022992
2	3.969757	-0.550075	-0.323569	0.397964	-0.322877	0.271493	0.076506	-0.355050	0.020116	-0.022675
3	3.596713	6.905070	0.792832	-0.604828	0.243176	-0.616970	-0.068051	-0.100163	-0.043481	-0.053456
4	3.151092	-1.358072	-1.862234	-0.185251	0.311342	0.090778	0.308087	0.099057	-0.026574	0.034113
...
564	4.436781	-2.020670	-0.457737	-0.472738	-0.802882	0.553317	0.904138	-0.155407	0.022121	-0.005704
565	2.874757	-2.395235	1.169465	0.073879	-0.745189	-0.100192	0.355690	-0.158585	-0.037307	-0.017740
566	0.649583	-2.041783	1.710966	-0.271559	-0.101971	-0.040732	-0.188382	0.018907	-0.047129	0.000593
567	6.905664	1.216243	1.640120	0.153472	0.405966	0.468304	-0.309367	0.223487	0.004156	-0.018134
568	-3.970833	-0.989971	2.158277	0.205643	1.595720	0.480961	-0.026477	-0.527614	0.444687	-0.029325

[569 rows x 10 columns]

Aplicando `pca.components_`

Imagen 19. Paso 2 y 3.

En el paso 4, se decide el número de componentes principales, el cual tiene 3 opciones.

La primera opción es calcular el porcentaje de relevancia, es decir, entre el 75 y 90% de varianza total.

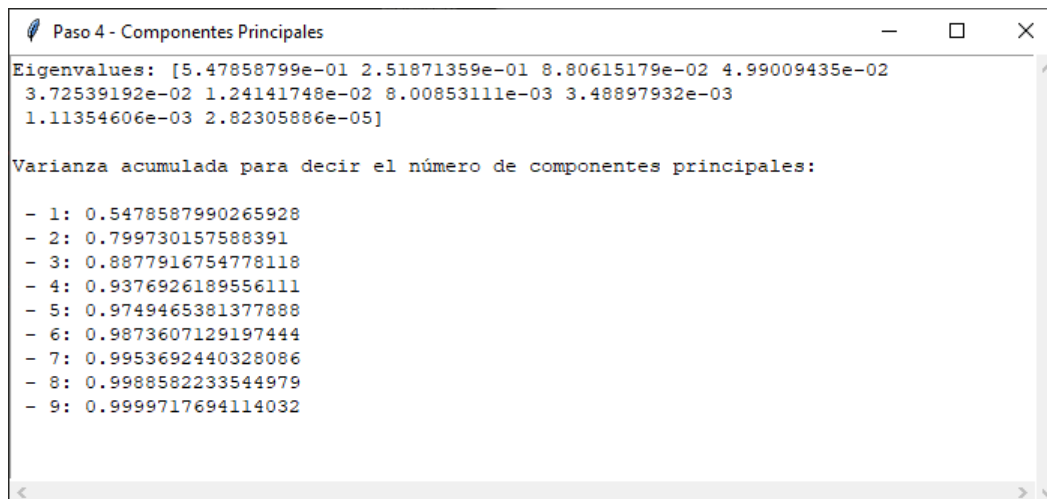


Imagen 20. Opción 1 del paso 4.

La segunda opción es la identificar, mediante una gráfica, el grupo de componente con mayor varianza.

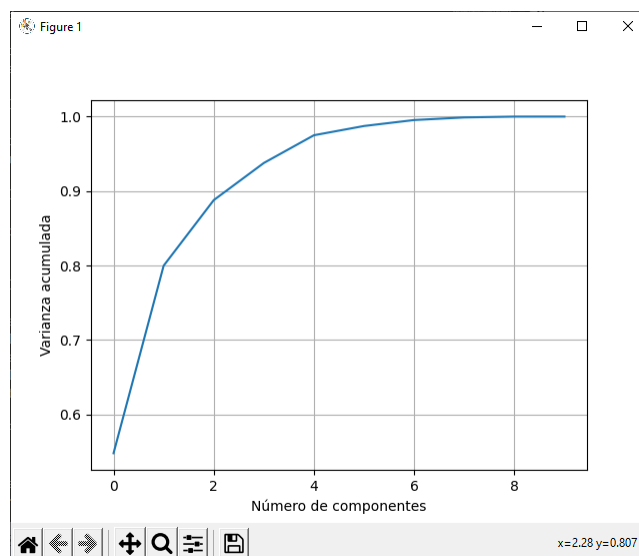


Imagen 21. Opción 2 del paso 4.

La última opción es elegir las dimensiones cuya varianza sea mayor a 1.

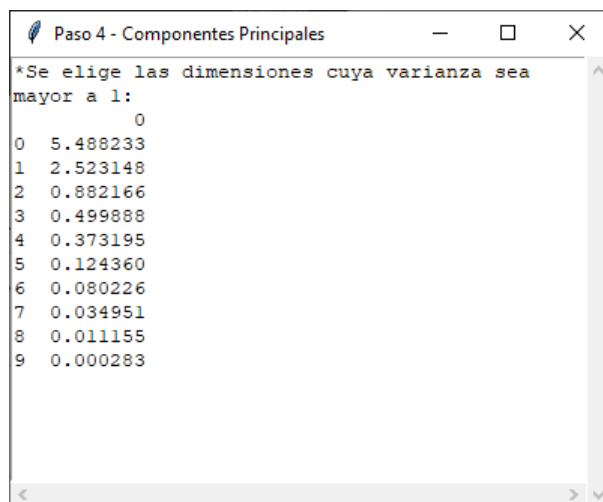


Imagen 22. Opción 3 del paso 4.

El último paso es examinar la proporción de relevancias -cargas-. El usuario lo que va a hacer es revisar los valores absolutos de los componentes principales seleccionados. Cuanto mayor sea el valor absoluto, más importante es esa variable en el componente principal.

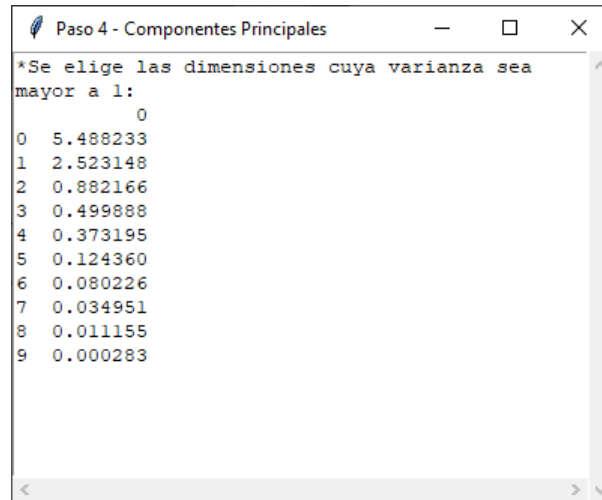


Imagen 23. Paso 5, selección de componentes principales con la ayuda de la varianza.

Para el siguiente algoritmo, el cual es el cluster particional, se muestran las instrucciones a seguir, así como la lista de las variables. Ahora lo que se va a hacer es seleccionar aquellas que se quieran agrupar, las cuales deben de ser las que tengan valores numéricos. Teniendo en cuenta que, al escribir una variable, se debe de dar click en Agregar. Cuando se hayan ingresado dos o más variables para poder realizar la agrupación, lo que se debe de hacer es dar click en Aceptar. Si se desea eliminar las variables, lo único que se debe de hacer es pulsar en Borrar Lista. Después aparecerá el botón de Ver Elbow Method en la parte inferior derecha como se ve en la imagen 24. Posterior a poner un valor para poder realizar lo que sigue del algoritmo, se presentarán otros dos botones para poder visualizar los resultados finales.

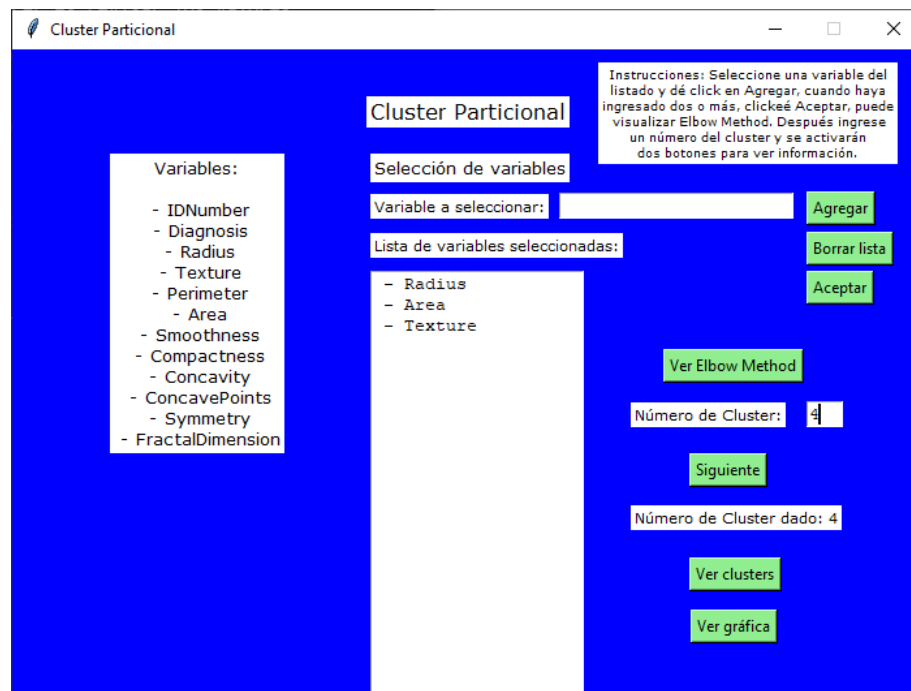


Imagen 24. Ventana de Cluster Particional junto con el proceso que se llevó a cabo.

En el botón Ver Elbow Method, se muestra el método del codo en donde se tiene que ver un doblez de la gráfica, ese valor es la que hay que considerar para el número de cluster, pero dependerá de la percepción de cada persona.

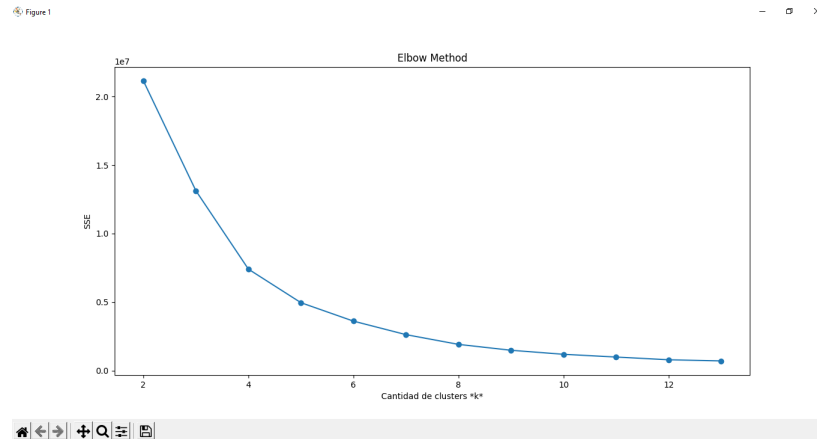


Imagen 25. Elbow Method de los datos que se eligieron.

En el botón Ver clusters se va a mostrar los clusters con la cantidad que tiene cada una de ellas.

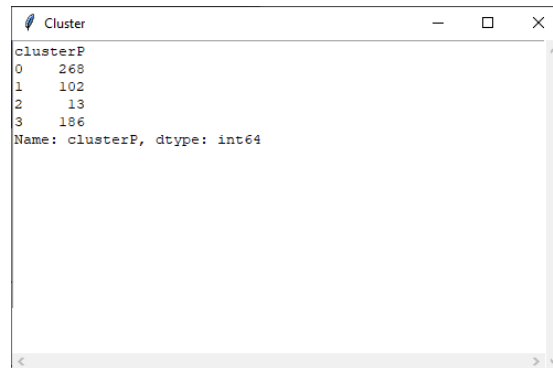


Imagen 26. Números de datos en cada cluster.

En el botón Ver Gráfica se puede observar de mejor manera cómo es que están agrupados los clusters.

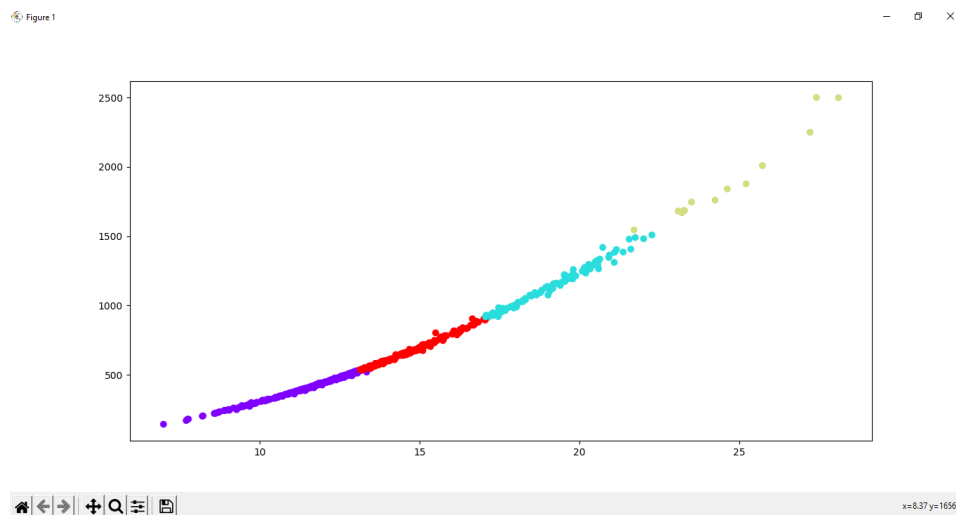


Imagen 27. Gráfica de los clusters.

Para finalizar, se tiene la Regresión Logística, la cual va a mostrar una pantalla en la cual se debe de elegir el porcentaje para la prueba y el seed, el cual sirve para inicializar el generador de números

aleatorios (Debe ser mayor que 0, en caso contrario, se dará el valor por defecto, que es 1234). Si no se da alguno de estos valores, el programa va a mandar un mensaje el cual dirá que se debe de escribir el seed o escoger el porcentaje de prueba.

Cuando se da estos valores y se da click en el botón Ver exactitud, se muestra dos botones, en uno de ellos se puede desplegar la información con respecto a la precisión de acertar con el pronóstico, en otro se va a confirmar que se utilizará ese valor de exactitud. Además, también se imprime el resultado de la exactitud.



Imagen 28. Ventana de Regresión Logística.

Al momento de pulsar Información se desplegará lo siguiente.

	precision	recall	f1-score	si
0	0.86	0.76	0.81	
1	0.87	0.93	0.90	
accuracy			0.87	
macro avg	0.87	0.85	0.85	
weighted avg	0.87	0.87	0.87	

Imagen 29. Ventana de la información sobre la exactitud y precisión del pronóstico.

Al darle Confirmar, se va a abrir una nueva ventana, en la cual se deberán ingresar los valores en cada uno de los espacios. Cuando se tengan llenados todos los espacios en blanco, a excepción del ID, se pulsa en Obtener resultado y se imprimirá el diagnóstico.

Poniendo un poco en contexto esta parte, un doctor con este programa llega a pronosticar si un paciente que tiene un tumor, con los valores de las variables, es maligno o benigno.

Imagen 30. Diagnóstico con los valores que se dieron.

6. Conclusiones

En este reporte, por cuestiones del número de páginas, únicamente se probó con un archivo; si se seleccionan con otros documentos que sean compatibles con los algoritmos que se presentan en dicho proyecto, no se tendrá problema alguno si se carga otro.

No fue sencillo la implementación, ya que se tuvo que pasar por distintos caminos para llegar a los resultados que se presentan en este proyecto. Respecto a lo que se comentó previamente, no se tenía algún conocimiento avanzado sobre el desarrollo web y se optó por algo que se conocía con anterioridad con el lenguaje de programación Python con la librería Tkinter.

Si al proyecto se le da un mejor desarrollo y estructura tanto interna como externa, puede haber una gran cantidad de modelos que se utilizarían para distintas áreas y que, sin duda alguna, serían de bastante ayuda para cualquier comunidad. Afortunadamente se realizaron los algoritmos que se requerían.

La realización de la Interfaz Gráfica de Usuarios ha sido de un gran esfuerzo y que sirvió para entender de una mejor manera los algoritmos que se llevaron a cabo durante este proyecto. Además, se obtuvo un crecimiento personal, ya que este proyecto aportó a tener optimas características con respecto a las capacidades de cómo resolver los problemas, así como el aumento de habilidades en la programación.

Referencias

- [1] Dr. Guillermo Gilberto Molero Castillo. (2021) Apuntes del curso de Minería de Datos
- [2] CódigoFacilito. (s. f.). ¿Por qué aprender Flask?. Recuperado 8 de agosto de 2021, de <https://codigofacilito.com/articulos/por-que-flask>
- [3] Fernandez, R. (2021, 4 enero). Tkinter - Introducción. Recuperado 8 de agosto de 2021, de <https://unipython.com/tkinter-introduccion/>
- [4] Alberca, A. S. (2021, 14 mayo). La librería Pandas. Recuperado 8 de agosto de 2021, de <https://aprendeconalf.es/docencia/python/manual/pandas/>
- [5] Alberca, A. S. (2020, 4 octubre). La librería Numpy. Recuperado 8 de agosto de 2021, de <https://aprendeconalf.es/docencia/python/manual/numpy/>
- [6] Matplotlib. (s. f.). Pyplot tutorial — Matplotlib 3.4.2 documentation. Recuperado 8 de agosto de 2021, de <https://matplotlib.org/stable/tutorials/introductory/pytest.html>
- [7] Katari, K. (2020, 11 agosto). Seaborn: Python - Towards Data Science. Medium. <https://towardsdatascience.com/seaborn-python-8563c3d0ad41>

- [8] Kite. (s. f.). Code Faster with Line-of-Code Completions, Cloudless Processing. Recuperado 8 de agosto de 2021, de <https://www.kite.com/python/docs/sklearn>
- [9] JMP. (s. f.). Análisis exploratorio de datos. Recuperado 8 de agosto de 2021, de https://www.jmp.com/es_cl/statistics-knowledge-portal/exploratory-data-analysis.html