

RASEMUS v. 2.0

Manual de Usuario

Aquí se encuentran todos los pasos necesarios para la configuración y realización de tareas que tiene la herramienta.

GRIAS KDD
01/01/2018

RASEMUS V 2.0

Rasemus es una herramienta débilmente acoplada con el sistema gestor de base de datos PostgreSQL con Licencia Pública General de GNU (GPL) orientada principalmente a proteger la libre distribución, modificación y uso. Desarrollada por el grupo de investigación GRIAS (Grupo de Investigación Aplicado a Sistemas), del programa de ingeniería de sistemas de la Universidad de Nariño de la ciudad de San Juan de Pasto (Nariño - Colombia).

La palabra Rasemus es una palabra del latín que significa racimo o grupo.

Rasemus es un software de apoyo a la toma de decisiones, mediante la ejecución de algoritmos de clustering.

Rasemus le permite al usuario escoger diferentes fuentes de datos como lo son archivos con formato .ARFF, formato .XLS, archivos planos y el sistema gestor de base de datos PostgreSQL. Permite utilizar filtros para la limpieza, selección y transformación del conjunto de datos. Rasemus le da la posibilidad de utilizar 3 técnicas de clustering que son: la técnica de clustering particional, con los algoritmos kmeans, kprototype y kfrequency, la técnica de clustering jerárquico con tres tipos de enlace: enlace mínimo (Single-link), enlace promedio (Averange-link) y enlace Máximo (Complete-link) y la técnica de clustering basada en densidad con el algoritmo DBScan con sus variaciones barrido directo () y barrido conectado. La presentación de los resultados se puede ver informes textuales que resumen todo el comportamiento de los algoritmos permitiendo su exportación a formatos .PDF, .HTML y .RTF, también utilizando graficas de dispersión para la visualización de los segmentos formados y la visualización del dendograma resultado de los algoritmos jerárquicos y la posibilidad de exportar las estas imágenes.

Su interfaz gráfica es amigable, diseñada con la técnica Drag & Drop para la realización del flujo de trabajo, permitiendo a los usuarios construir el proceso de selección de las operaciones necesarias de forma ágil. Además su manejo de colores hace que Rasemus tenga una presentación moderna de sus iconos, acorde a la tendencia actual.

REQUERIMIENTOS

Lo único que se necesita para el correcto funcionamiento de la herramienta en cualquier sistema operativo, es tener instalado la máquina virtual de Java 1.6 o superior.

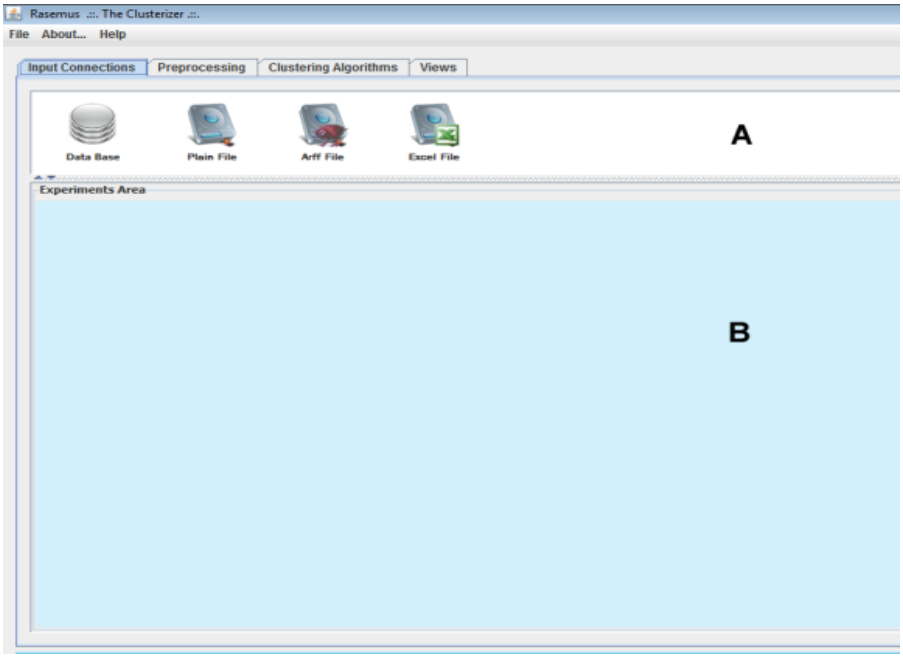
Nota: En las versiones anteriores de la máquina virtual pueden perderse algunos elementos.

1. EMPEZAR A UTILIZAR RASEMUS

Rasemus es desarrollo bajo el lenguaje de programación Java, lo que le permite ser multiplataforma, para ello Rasemus se presenta en un archivo .jar. para no perder la opción de ser una herramienta multiplataforma. Se toma el archivo .jar y se da doble click.

La ventana principal está compuesta por dos sectores. En la figura 1 se muestra la ventana principal en donde se identifican dos zonas. La zona A es la parte en donde se presentan las tareas que se necesitan para realizar las tareas de descubrimiento de conocimiento, representadas en iconos, divididos en pestañas según su especialidad. En la parte B, está el área de trabajo, en donde se colocan los iconos que el usuario elija trabajar.

Figura 1: Ventana principal

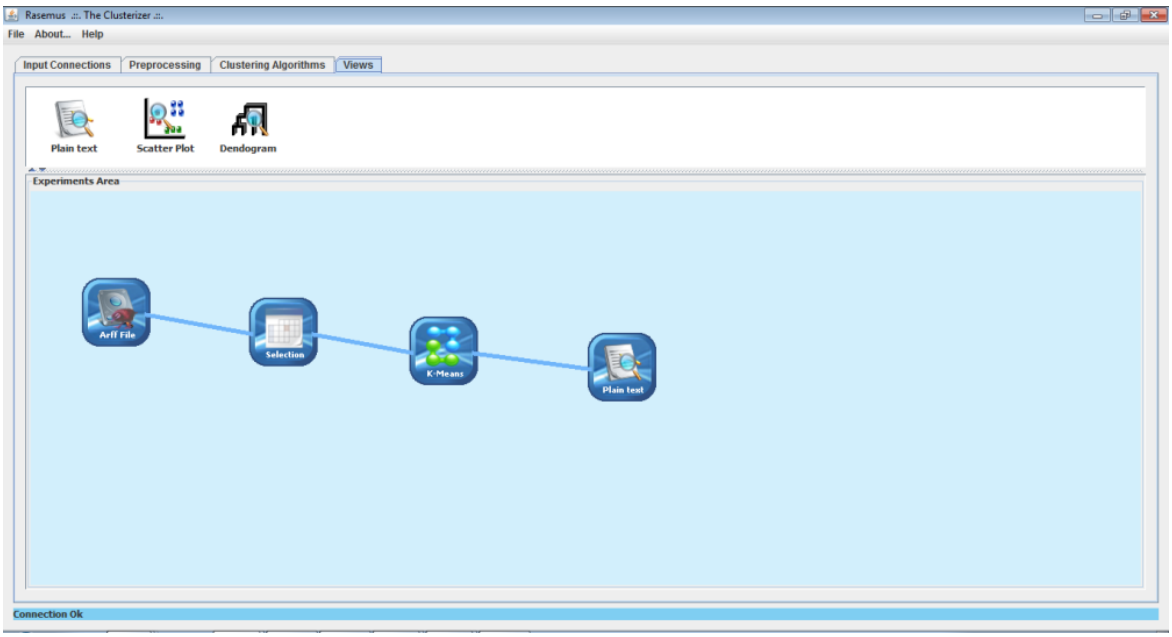


Para la colocación de los iconos de las tareas en el área de trabajo se debe dar click sostenido sobre el icono y arrástralo hasta el lugar del área de trabajo donde desee colocar el icono. Para unir los iconos y crear flujos de trabajos, los iconos están equiparados con dos áreas de conexión, una en la parte derecha, y otra en la parte izquierda, que funcionan de la siguiente manera: primero se da click sostenido en el área del icono que se decida como fuente, sin soltar el botón derecho del mouse, se arrastra el puntero hasta el área de conexión del icono destino y se suelta el botón y se establece la conexión.

Figura 2: Forma de Conexión



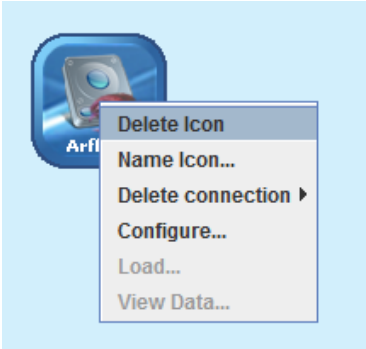
Figura 3: Manejo de la operaciones.



Los iconos ubicados en el área de trabajo son configurados, colocando el puntero del mouse sobre el icono y dando click derecho. Ahí se despliega un menú en donde se presentan las diferentes opciones de configuración dependiendo del tipo de icono seleccionado (ver figura 4). Los iconos tienen como opciones básicas:

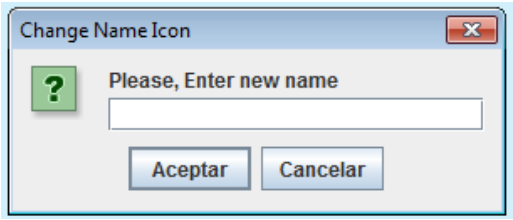
Delete icon
Permite borrar el icono

Figura 4: Icono de Rasemus.



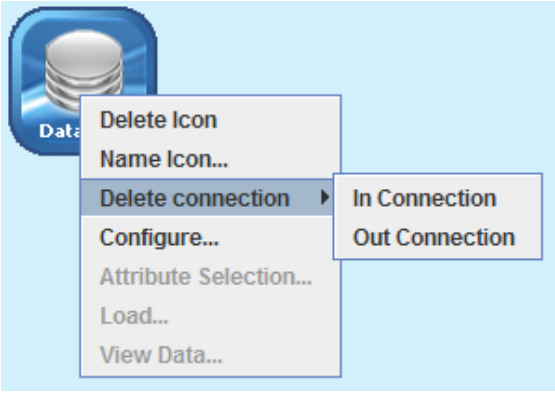
Name Icon
Permite cambiar el nombre del icono, desplegando una ventana emergente (ver figura 5)

Figura 5: Opción Name Icon.



Name Icon
Permite borrar las conexiones del icono, esta opción despliega un submenú en donde se puede establecer qué tipo de conexiones quiere borrar, las entrantes con *In Connections* o las salientes con *Out Connections.* Ver figura 6.

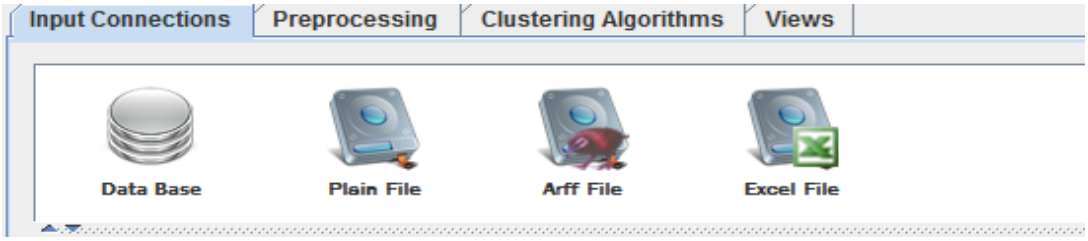
Figura 6: Opción Delete connections.



1.1 Input Connections o Selección Fuente de Datos:

En la pestaña Input Connections, se presentan los diferentes iconos que permiten la selección de fuentes de datos, entre ellas está el icono Data Base, que permite la conexión con el sistema gestor de bases de datos PostgreSQL, el icono Plain File, que permite la conexión con tipos de datos con formato plano, también se encuentra el icono Arff File que permite la conexión con archivos de formato .Arff y por ultimo esta el icono Excel File, que permite la conexión con archivos con formato .Xls. ver figura 7.

Figura 7: Iconos de la opción Input Connections.

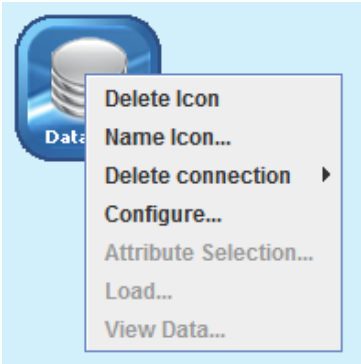


1.1.1 Icono Data Base



Permite la conexión con el sistema gestor de bases de datos PostgreSQL, con las opciones que se presenta en la figura 8.

Figura 8: Opciones Icono Data Base.

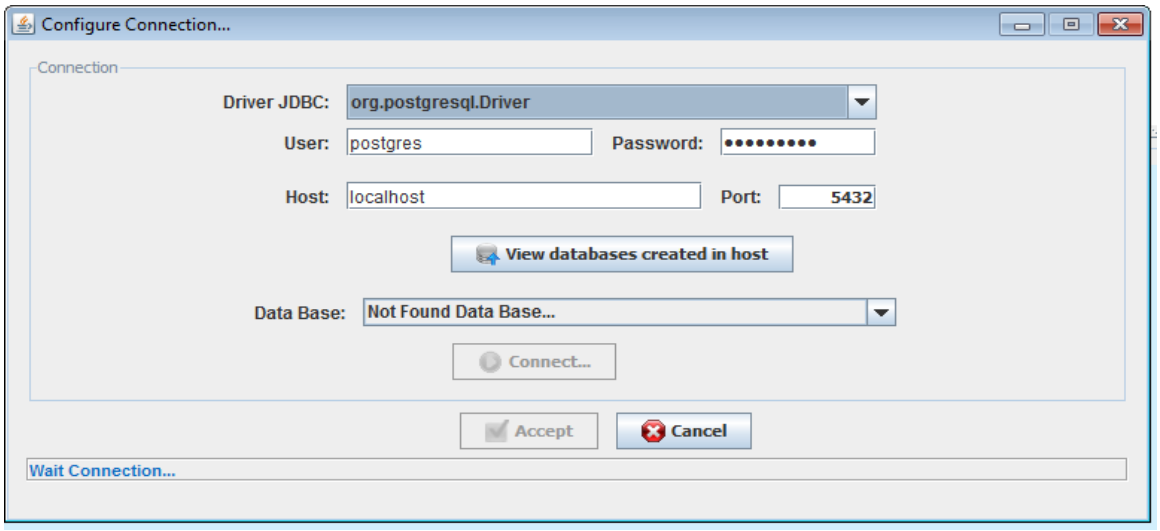


Configure

La opción configure despliega una ventana que solicita los parámetros básicos para la establecer la conexión a una base de datos (ver figura 9), como el usuario, el password, el servidor y el puerto. Ya configurados estos elementos se presiona el botón *View Found created in host,* que lista las

bases de datos creadas en el servidor, permitiéndole al usuario escoger la base de datos que desea trabajar. Por último se presiona el botón *Connect* y si los datos son correctos, en la barra de estado que está en la parte inferior presenta un mensaje OK; si los datos suministrados no son correctos se presenta el mensaje de error correspondiente. Si la conexión es correcta, para finalizar se presiona el botón *Accept*. Se habilita la opción configure.

Figura 9: Ventana de configuración de conexión a base de datos.



***Attribute Selection*:**

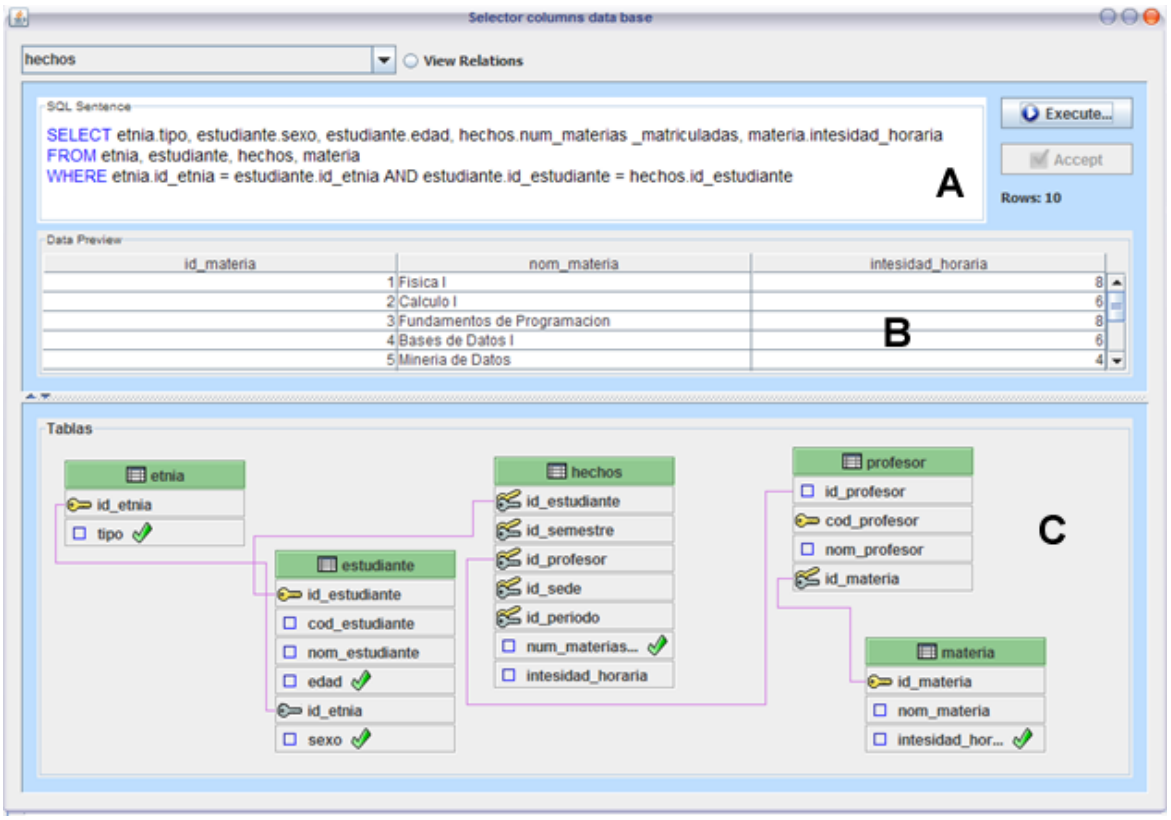
La opción *Attribute selection* permite seleccionar los atributos de las tablas que se encuentran en a base de datos, las tablas de la base de datos se presentan en un menú desplegable ubicado en la parte superior de la ventana, el usuario elije la tabla a trabajar y se adiciona al área de trabajo. la ventana se encuentra separada en tres partes (ver figura 10):

A) área de visualización del sql, en donde se va mostrando la construcción de la sentencia sql, con los diferentes atributo seleccionados. Aquí se encuentran los botones *Execute* que permite ejecutar la sentencia sql formado, y hace visualizar lo resultados en la parte B.

B) en esta zona se presenta el resultado de la sentencia sql formada.

C) En esta zona se encuentran las tablas seleccionadas por el usuario. Presenta las tablas como objetos independientes dando la posibilidad que con click sostenido sobre su cabecera pueda ubicarla en cualquier lugar de su área. Para selección el atributo que se desea trabajar se debe dar click sobre la parte izquierda del atributo. En el área de trabajo también observan las llaves primarias y foráneas de las tablas, permitiendo también ver una línea de relación entre las referencias de las tablas que facilitan la labor de la selección de atributos.

Figura 10: Ventana de selección de atributos.



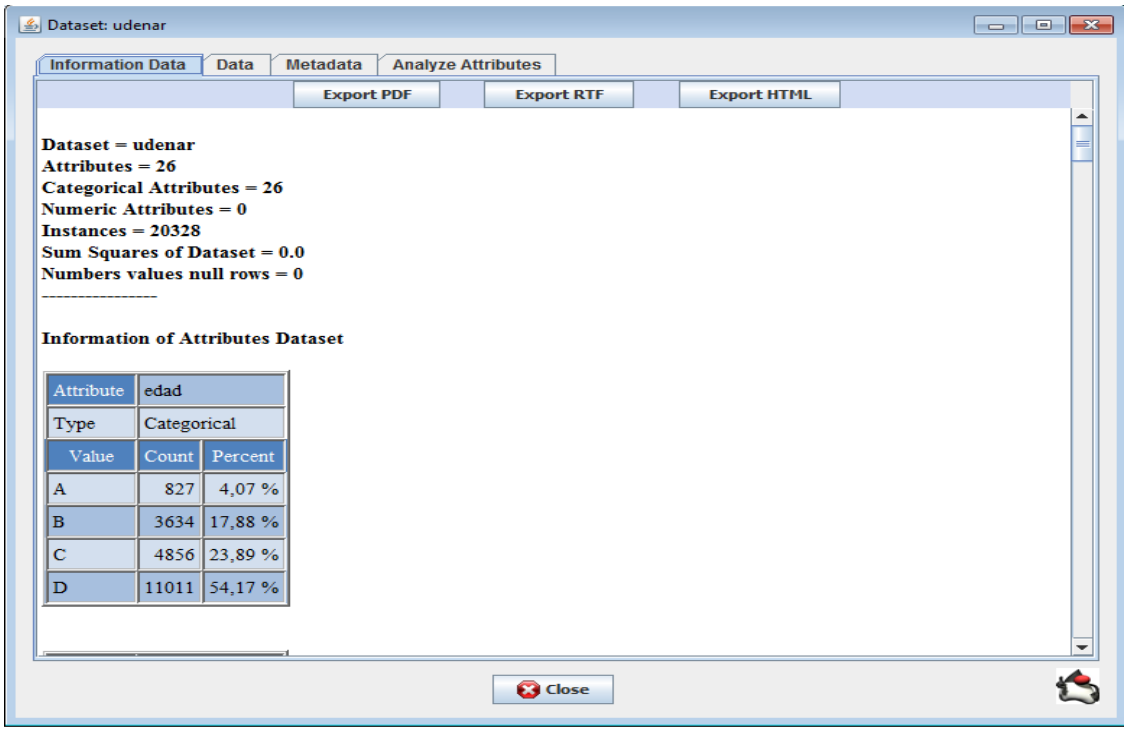
Load

Permite cargar los datos de los atributos seleccionados a memoria. Sin esta opción no se puede conectar el icono con ningún otro icono.

View Data

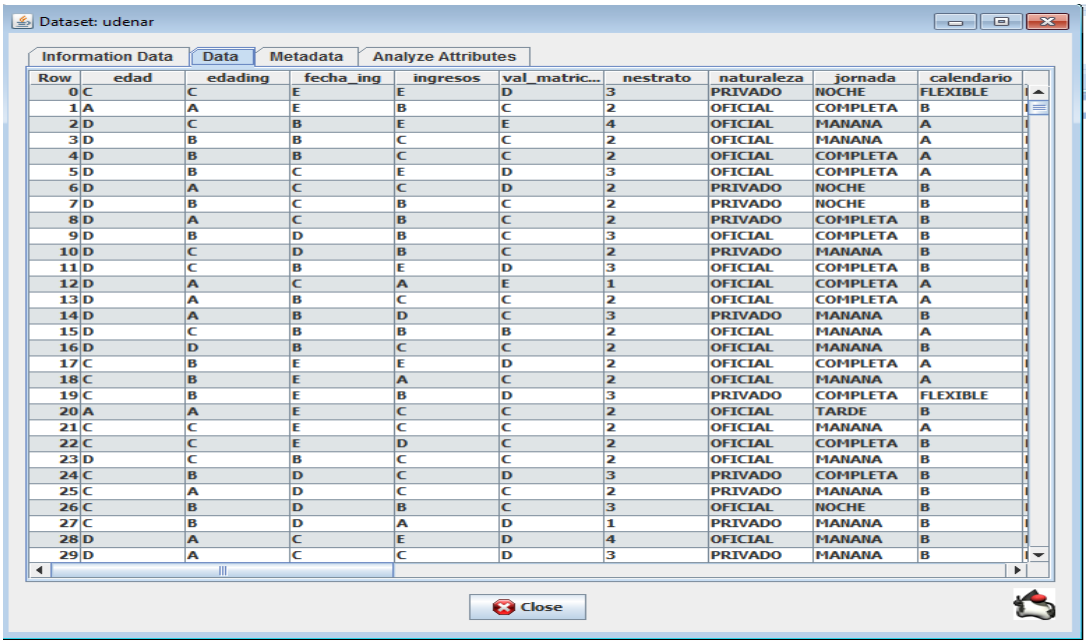
Permite mirar información acerca del conjunto de datos. En la primera pestaña *Information Data* se presenta las características del conjunto de datos y sus atributos, esta información se puede exportar a archivos con formato .PDf, .RTF y .HTML.(ver figura 11)

Figura 11: Pestaña *Information Data* opción View Data.



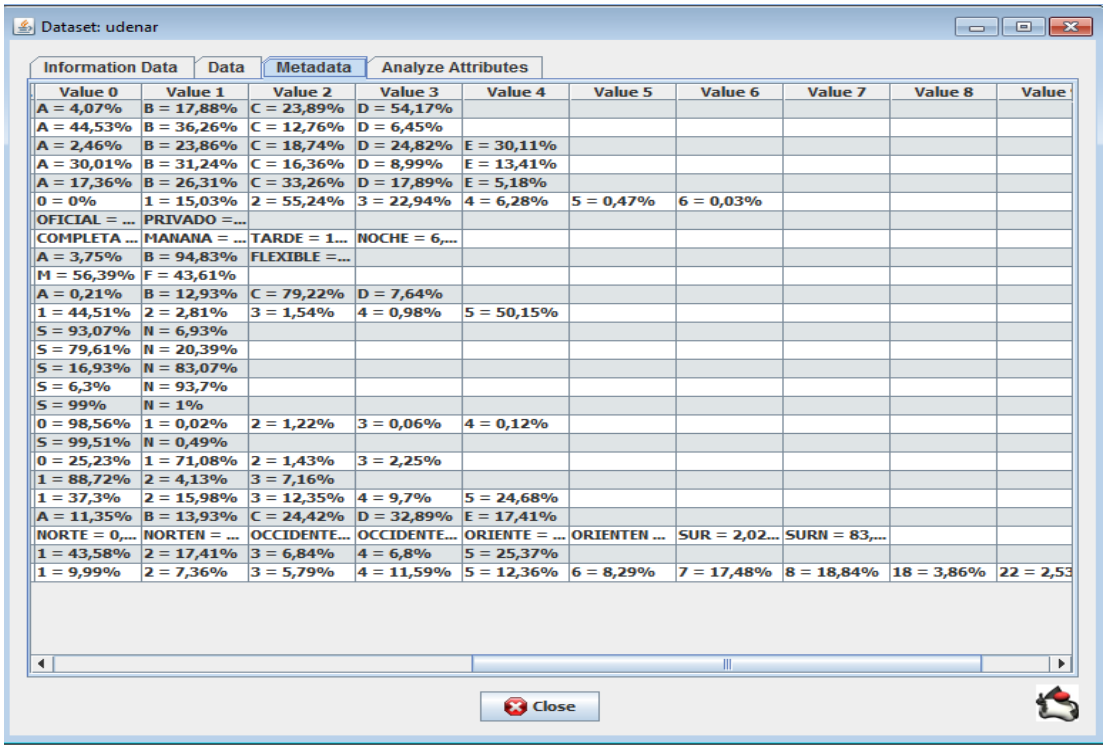
La segunda pestaña denominada *Data* presenta los registros del conjunto de datos. Aquí se puede ordenar el por atributo los datos dando click sobre el nombre de la columna. Ver figura 12.

Figura 12: Pestaña *Data* opción View Data.



La tercera pestaña denominada *Metadata* presenta las características de los atributos, para el caso de los atributos presenta el valor mínimo, la media, el valor máximo y mucho mas, para los atributos de tipo categórico se presenta sus valores con sus respectivas frecuencias absolutas y porcentaje en el conjunto de datos. Esta información le permite comprender mejor sus datos de sus datos, ver figura 13.

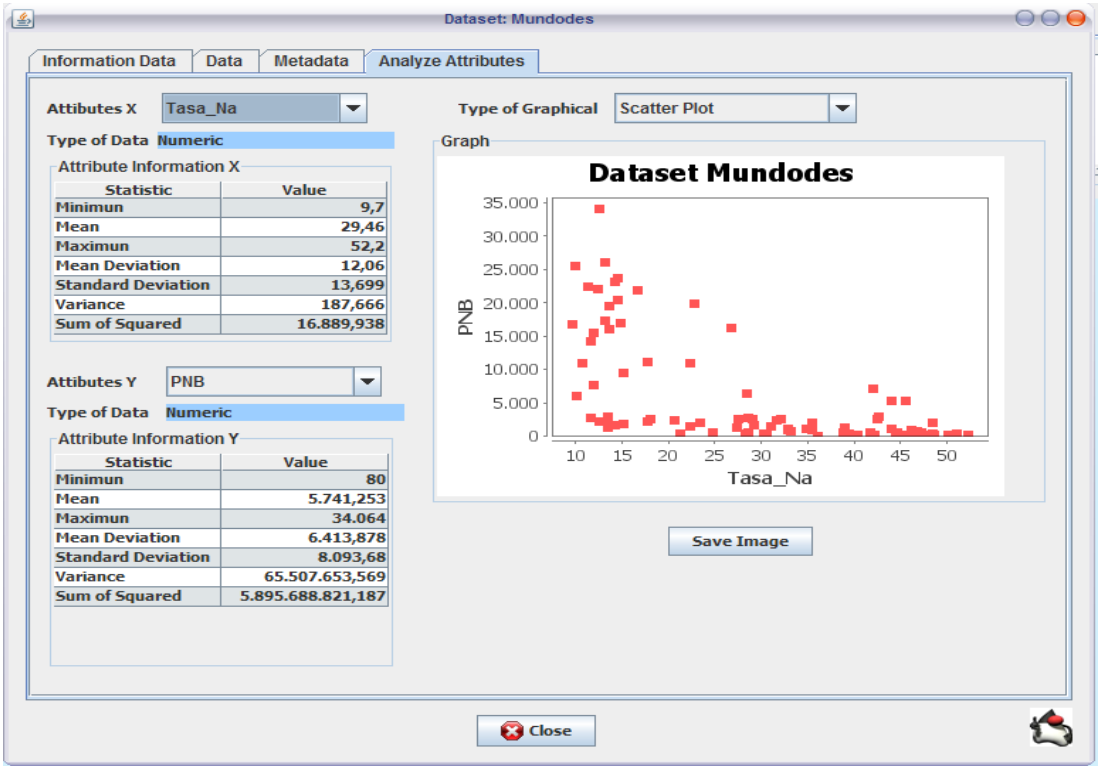
Figura 13: pestaña *Metadata* opción View Data.



Value 0	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8	Value
A = 4,07%	B = 17,88%	C = 23,89%	D = 54,17%						
A = 44,53%	B = 36,26%	C = 12,76%	D = 6,45%						
A = 2,46%	B = 23,86%	C = 18,74%	D = 24,82%	E = 30,11%					
A = 30,01%	B = 31,24%	C = 16,36%	D = 8,99%	E = 13,41%					
A = 17,36%	B = 26,31%	C = 33,26%	D = 17,89%	E = 5,18%					
0 = 0%	1 = 15,03%	2 = 55,24%	3 = 22,94%	4 = 6,28%	5 = 0,47%	6 = 0,03%			
OFICIAL = ...	PRIVADO =...								
COMPLETA ...	MANANA = ...	TARDE = 1...	NOCHE = 6,...						
A = 3,75%	B = 94,83%	FLEXIBLE =...							
M = 56,39%	F = 43,61%								
A = 0,21%	B = 12,93%	C = 79,22%	D = 7,64%						
1 = 44,51%	2 = 2,81%	3 = 1,54%	4 = 0,98%	5 = 50,15%					
S = 93,07%	N = 6,93%								
S = 79,61%	N = 20,39%								
S = 16,93%	N = 83,07%								
S = 6,3%	N = 93,7%								
S = 99%	N = 1%								
0 = 98,56%	1 = 0,02%	2 = 1,22%	3 = 0,06%	4 = 0,12%					
S = 99,51%	N = 0,49%								
0 = 25,23%	1 = 71,08%	2 = 1,43%	3 = 2,25%						
1 = 88,72%	2 = 4,13%	3 = 7,16%							
1 = 37,3%	2 = 15,98%	3 = 12,35%	4 = 9,7%	5 = 24,68%					
A = 11,35%	B = 13,93%	C = 24,42%	D = 32,89%	E = 17,41%					
NORTE = 0,...	NORTEN = ...	OCCIDENTE...	OCCIDENTE...	ORIENTE = ...	ORIENTEN ...	SUR = 2,02...	SURN = 83,...		
1 = 43,58%	2 = 17,41%	3 = 6,84%	4 = 6,8%	5 = 25,37%					
1 = 9,99%	2 = 7,36%	3 = 5,79%	4 = 11,59%	5 = 12,36%	6 = 8,29%	7 = 17,48%	8 = 18,84%	18 = 3,86%	22 = 2,53

La cuarta pestaña denominada *Analyze attributes* presenta las características de los atributos. Dando la opción de crear graficas de diferente tipo como diagramas de dispersión, si es atributo numérico o diagramas de pastel si son categóricos y demás graficas. Estas graficas pueden ser exportadas como JPG, presionando el botón *save image*. Ver figura 14.

Figura 14: pestaña *Analyze attributes* opción View Data.

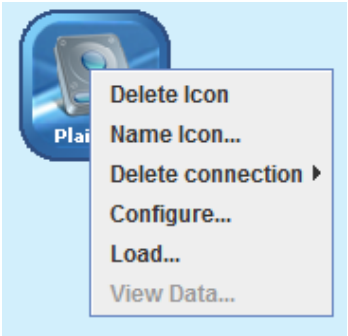


1.1.2 Icono Plain File



Permite la conexión a archivos separados por COMAS, PUNTO Y COMA, TAB, espacios u otros tipos de separadores. En la figura 15 se presentan las opciones de este icono.

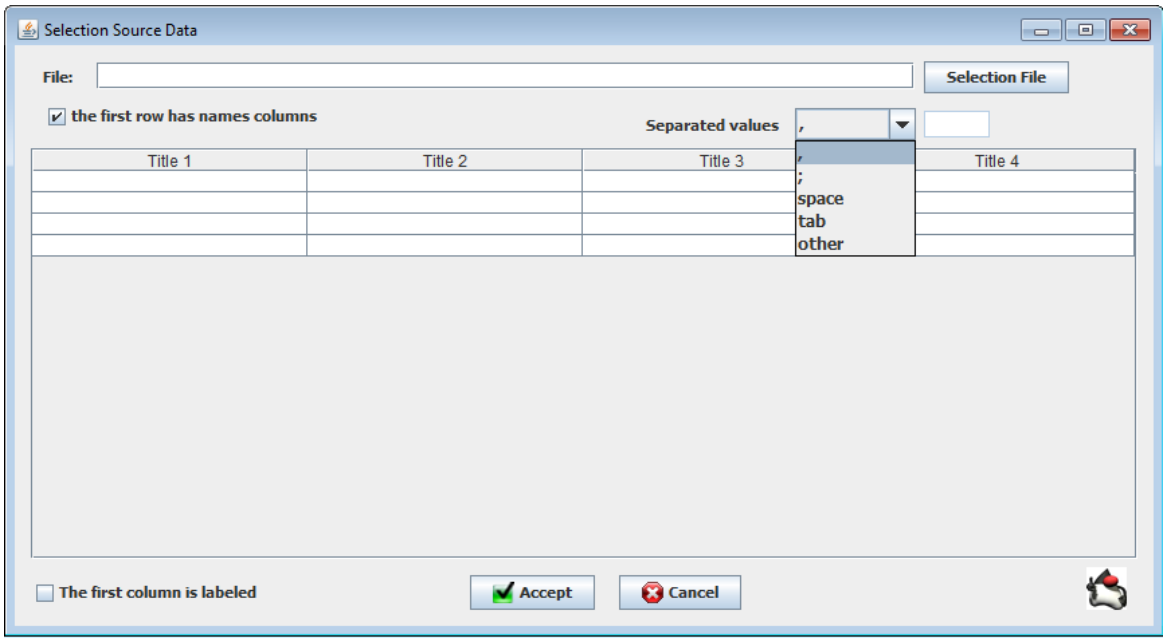
Figura 15: Opciones Icono Plain File.



Configure

Despliega una ventana que permite seleccionar el archivo, presionando en el botón *Selection File*, en esta ventana también tiene la opción de establecer si se quiere tomar la primera fila como el nombre de la columnas seleccionando *the firts row has names columns*. Otra opción es seleccionar *the firts columns is labeled*, que permite tomar la primera columna como etiquetas o identificadores de los registros para su identificación. Si todos la configuración es correcta se habilita la opción *load*. También aquí se presenta una previsualización de los datos reconocidos.

Figura 16: Ventana de configuración Plain File.



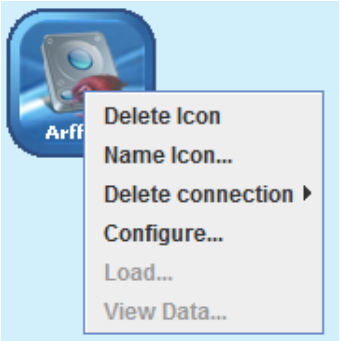
Load
Permite cargar los datos del archivo a memoria. Sin esta opción no se puede conectar el icono con ningún otro icono.

View Data
Es igual a la opción del icono Data Base.

1.1.3 Icono Arff File



Permite la conexión a archivos con formato .ARFF que son archivos creados por la herramienta Weka. Las opciones *Configure, load y View Data* son iguales a las del Icono Plain File.
Figura 17: Opciones Icono Arff File.

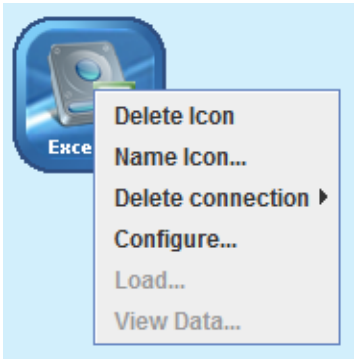


1.1.4 Icono Excel File



Permite la conexión a archivos con formato .XLS que son archivos creados, por Microsoft Excel. Las opciones *Configure, load y View Data* son iguales a las del Icono Plain File.

Figura 18: Opciones Icono Excel File.

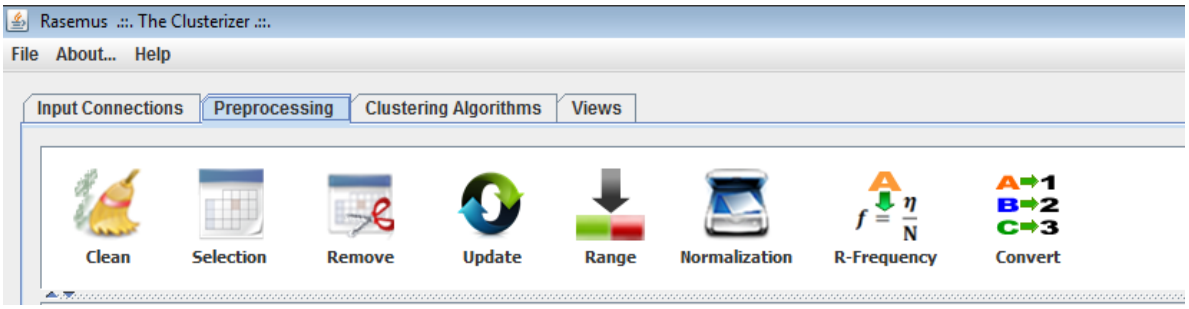


1.2 *Preprocessing* o Preprocesamiento de Datos:

En la pestaña Input *Preprocesing*, se presentan los diferentes iconos que permiten limpieza, selección y transformación del conjunto de datos, entre ellos están el icono *Clean* que permite eliminar los datos de tipo nulo, el Icono *selection* , que permite seleccionar algunos atributos del

conjunto de datos, el icono *Update* permite actualizar los valores de los atributos, el icono *Range* permite crear rangos en los atributos numéricos, el icono *Normalization* permite normalizar los atributos de tipo numérico, el icono *R-Frequency* reemplaza los atributos categóricos por el valor de la frecuencia relativa de la categoría dentro del atributo y por ultimo esta el icono *Convert* que convierte los atributos categóricos a numéricos asignándole un número determinado o también sirve para transformar atributos numéricos a categóricos. Estos se presentan en la figura 19.

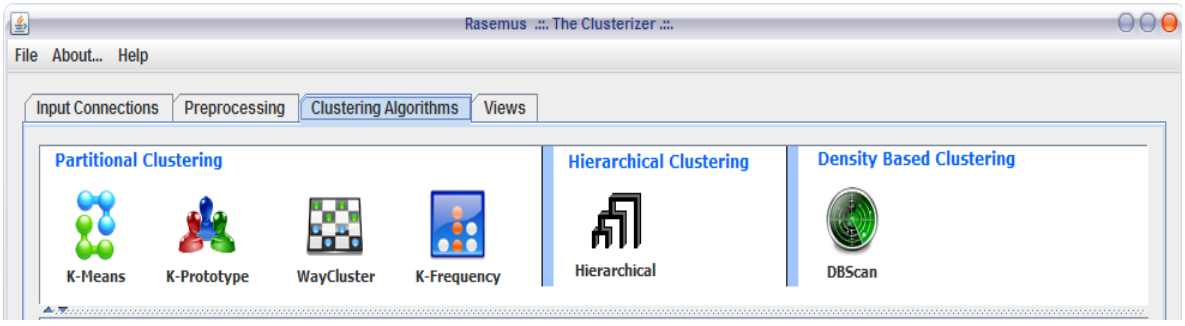
Figura 19: iconos de la opción *Preprocesing.*



1.3 *Clustering Algorithms* o Algoritmos de Clustering

Permiten realizar las tareas de clustering a un conjunto de datos. Estos algoritmos se encuentran divididos en tres tipos, los de tipo de clustering particional o *Partitional Clustering* con los algoritmos kmeans, kprototype y kfrequency, los algoritmos de clustering jerárquicos o *Hierarchical Clustering* y los algoritmos de clustering basado en densidad o *Density Based Clustering*. Para poder ejecutar este tipo de opciones los iconos deben tener conectado un Icono Data Base, Plain File, Arff File, Excel File y todos los atributos de preprocesamiento. (ver figura 20)

Figura 20: iconos de la opción Clustering Algorithms.

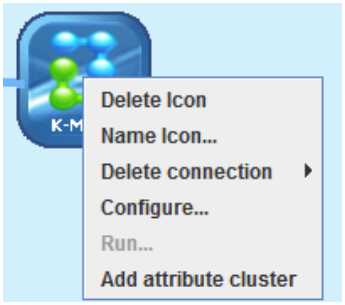


1.3.1 Icono K-Means



Crea segmentos del conjunto de datos a partir del algoritmo K-Means. En la figura 20 se muestras la opciones que presenta este icono (ver figura 21).

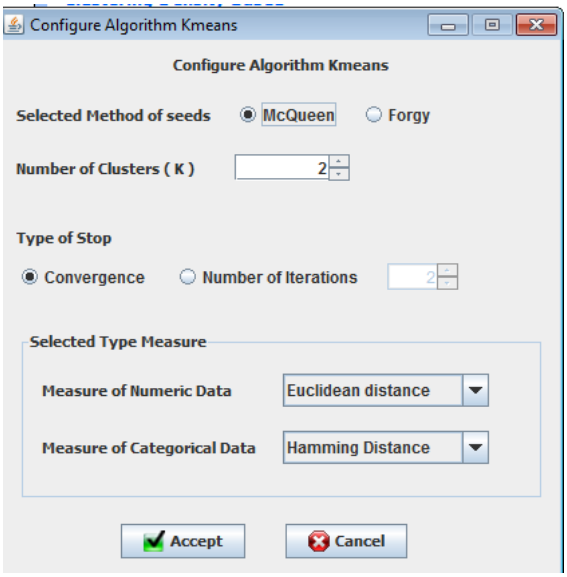
Figura 21: Opciones Icono K-Means.



Configure

Permite al usuario configurar el algoritmo kmeans, permitiéndole al usuario la opción de escoger el método de selección de semillas ya sea por McQueen que toma los primeros registros como semillas o la opción Forgy que toma semillas de forma aleatoria. La otra opción es definir el número de segmentos a formar. La siguiente es la opción de escoger el tipo de parada, entre las opciones están la convergencia, es decir cuando ya no haya cambios en los grupos la otra es según un determinado número de iteraciones. El siguiente parámetro permite elegir, qué formula de distancia se va a utilizar para los atributos numéricos y el otro que tipo de formula de similitud que se utilizara para los atributos categóricos, entre la formulas para tipos numéricos se tiene, la distancia euclidiana, la distancia Manhattan, distancia Chebyshev y mahalanobis y para los atributos categóricos se tiene las técnicas de similitud haming y proporción de coincidencias. Ver figura 22.

Figura 22: configuración icono kmeans



Run
Permite la ejecución de los algoritmos.

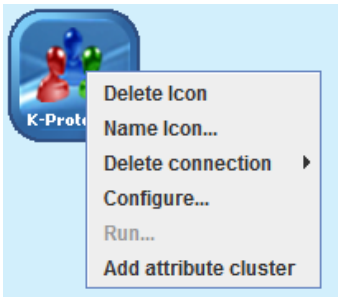
Add Attribute Cluster
Permite crear un archivo con formato .ARFF adicionando una columna en donde ira el numero del segmento al cual fue asignado el registro.

1.3.2 Icono K-Prototype



Crea segmentos del conjunto de datos a partir del algoritmo K-Prototype. En la figura 23 se muestran las opciones que presenta este icono.

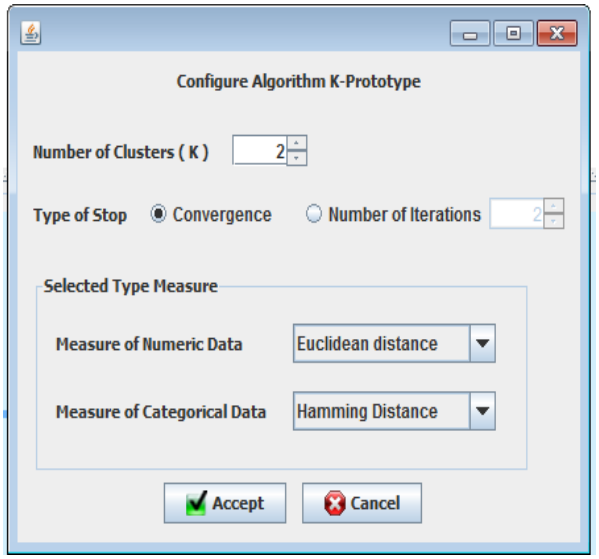
Figura 23: Opciones icono kprototype.



Configure

Permite al usuario configurar el algoritmo kprototype, permitiéndole al usuario la opción de definir el número de segmentos a formar. La siguiente es la opción de escoger el tipo de parada, entre las opciones están la convergencia, es decir cuando ya no haya cambios en los grupos la otra es según un determinado número de iteraciones. El siguiente parámetro permite elegir, qué formula de distancia se va a utilizar para los atributos numéricos y el otro que tipo de formula de similitud que se utilizara para los atributos categóricos, entre la formulas para tipos numéricos se tiene, la distancia euclidiana, la distancia Manhattan, distancia Chebyshev y mahalanobis y para los atributos categóricos se tiene las técnicas de similitud haming y proporción de coincidencias. Ver figura 24.

Figura 24: Configuración icono K-Prototype



Run
Permite la ejecución de los algoritmos.

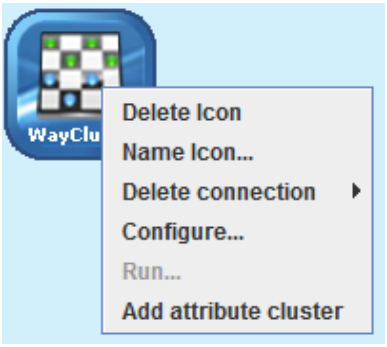
Add Attribute Cluster
Permite crear un archivo con formato .ARFF adicionando una columna en donde ira el numero del segmento al cual fue asignado el registro.

1.3.3 Icono wayCluster



Crea segmentos del conjunto de datos a partir del algoritmo wayCluster. En la figura 25 se muestran las opciones que presenta este icono.

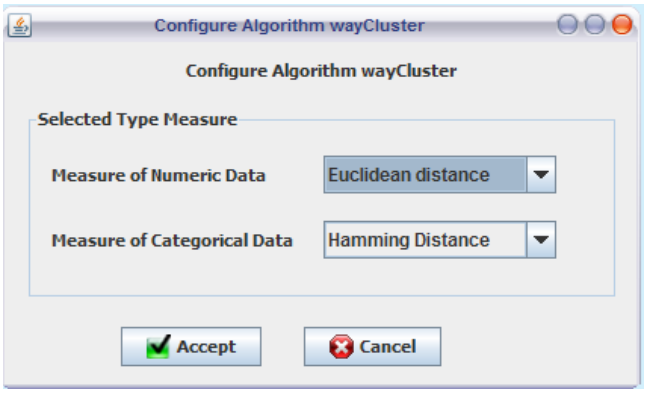
Figura 25: Opciones icono wayCluster.



Configure

Permite al usuario configurar el algoritmo wayCluster. Solo permite elegir qué formula de distancia se va a utilizar para los atributos numéricos y el otro que tipo de formula de similitud se utilizara para los atributos categóricos, entre la formulas para tipos numéricos se tiene, la distancia euclidiana, la distancia Manhattan, distancia Chebyshev y mahalanobis y para los atributos categóricos se tiene las técnicas de similitud haming y proporción de coincidencias. (ver figura 26)

Figura 26: Configuración icono wayCluster



Run
Permite la ejecución de los algoritmos.

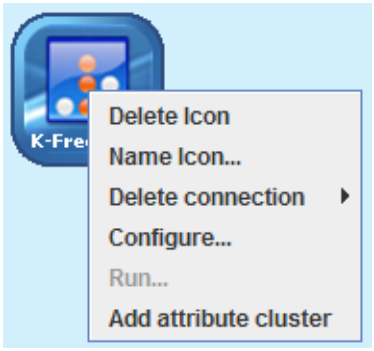
Add Attribute Cluster
Permite crear un archivo con formato .ARFF adicionando una columna en donde ira el numero del segmento al cual fue asignado el registro.

1.3.4 Icono K-Frequency



Crea segmentos del conjunto de datos a partir del algoritmo K-Frequency. En la figura 27 se muestran las opciones que presenta este icono.

Figura 27: Opciones icono K-Frequency



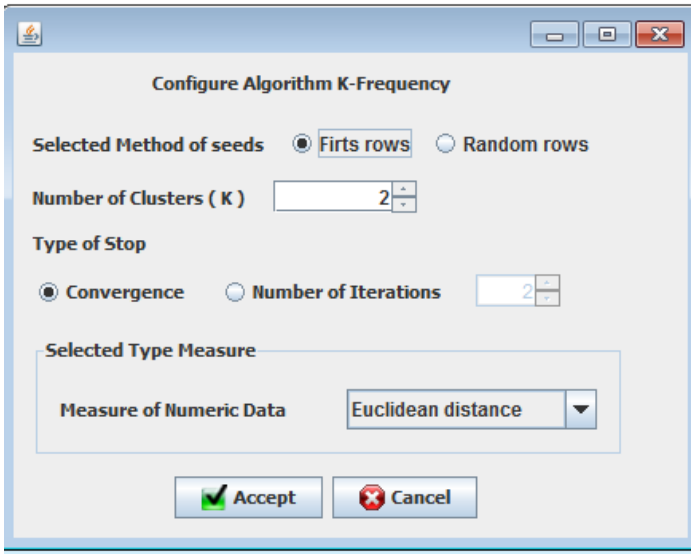
Configure

Permite al usuario configurar el algoritmo kfrequency, permitiéndole al usuario la opción de escoger el método de selección de semillas, utilizando los primeros registros como semillas utilizando la opción aleatoria que toma semillas aleatorias. La otra opción es definir el número de segmentos a formar. La siguiente es la opción de escoger el tipo de parada, entre las opciones están la convergencia, es decir cuando ya no haya cambios en los grupos o según un determinado número de iteraciones, el siguiente parámetro permite elegir qué formula de distancia se va a utilizar para los atributos numéricos entre las formulas se tiene la distancia euclidiana, la distancia Manhattan, distancia Chebyshev y mahalanobis y para los atributos categóricos se tiene las técnicas de similitud haming y proporción de coincidencias. Ver figura 28.

Run
Permite la ejecución de los algoritmos.

Add Attribute Cluster
Permite crear un archivo con formato .ARFF adicionando una columna en donde ira el numero del segmento al cual fue asignado el registro.

Figura 28: configuración icono K-Frequency.

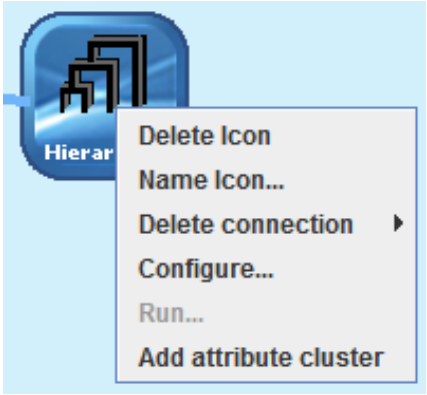


1.3.5 Icono Hierarchical



Crea segmentos del conjunto de datos a partir de algoritmos Jerárquicos de tipo aglomerativo. En la figura 29 se muestran la opciones que presenta este icono.

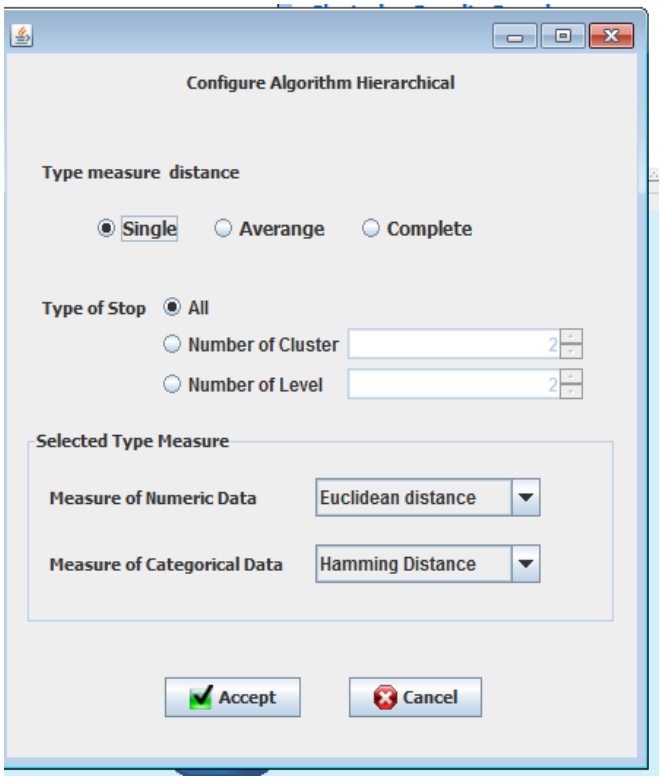
Figura 29: Opciones icono Hierarchical



Configure

Permite al usuario configurar que tipo de algoritmo jerárquico se desea utilizar, estos son: el algoritmo jerárquico de enlace mínimo, el cual corresponde a la opción *Single*, seguido de la opción de clustering jerárquico de enlace promedio que corresponde a la opción *Averange* y por último se tiene la opción complete que permite ejecutar el algoritmo con enlace máximo. La siguiente es la opción de escoger el tipo de parada, entre las opciones está formar un solo segmento o formar un determinado número de segmentos o alcanzar un cierto número de niveles o iteraciones en la creación del dendograma. , el siguiente parámetro permite elegir qué formula de distancia se va a utilizar para los atributos numéricos y el otro que tipo de formula de similitud se utilizara para los atributos categóricos, entre la formulas para tipos numéricos se tiene, la distancia euclidiana, la distancia Manhattan, distancia Chebyshev y mahalanobis y para los atributos categóricos se tiene las técnicas de similitud haming y proporción de coincidencias. Ver figura 30.

Figura 30: Configuración icono Hierarchical.



Run
Permite la ejecución de los algoritmos.

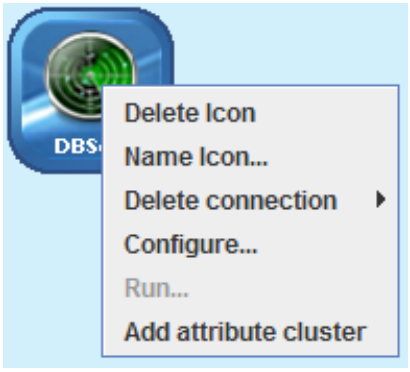
Add Attribute Cluster
Permite crear un archivo con formato .ARFF adicionando una columna en donde ira el numero del segmento al cual fue asignado el registro.

1.3.5 Icono DBScan



Crea segmentos del conjunto de datos a partir del algoritmo DBScan, que hace parte de los algoritmos de clustering basado en densidad. En la figura 31 se muestran la opciones que presenta este icono.

Figura 31: Opciones icono Hierarchical



Configure

Permite al usuario configurar los parámetros del algoritmo DBScan, estos son: el tamaño del área de barrido o el grado de densidad denominado épsilon. El siguiente parámetro es el número mínimo de puntos que debe haber en área de barrido para considerarse segmento. El tipo d barrido puede definir si va a ser solo el barrido directo o *Directly Density – Reachable* o la opción de barrido conectado o *Density - Connected* que permite la unión de elementos que estén cerca de los elementos segmentados. El siguiente parámetro permite elegir qué formula de distancia se va a utilizar para los atributos numéricos y el otro que tipo de formula de similitud se utilizara para los atributos categóricos, entre la formulas para tipos numéricos se tiene, la distancia euclidiana, la distancia Manhattan, distancia Chebyshev y mahalanobis y para los atributos categóricos se tiene las técnicas de similitud haming y proporción de coincidencias. Ver figura 32.

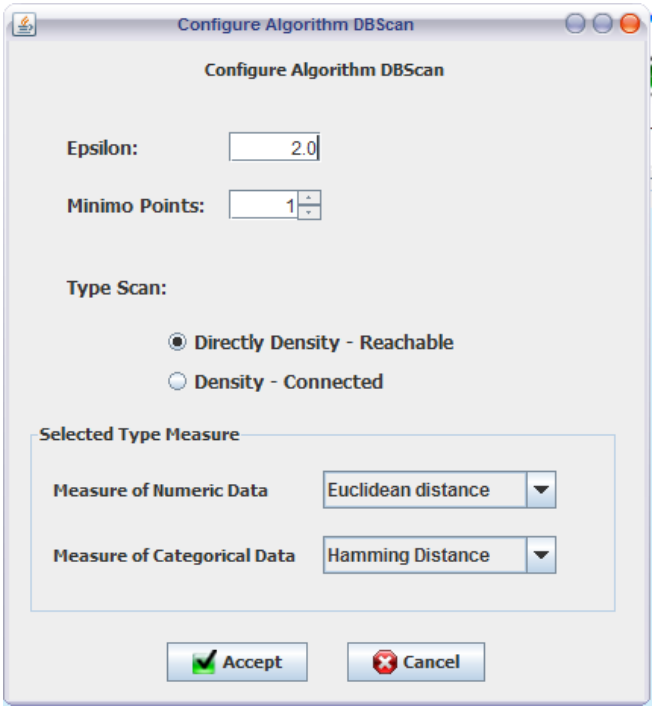
Run

Permite la ejecución de los algoritmos.

Add Attribute Cluster

Permite crear un archivo con formato .ARFF adicionando una columna en donde ira el numero del segmento al cual fue asignado el registro.

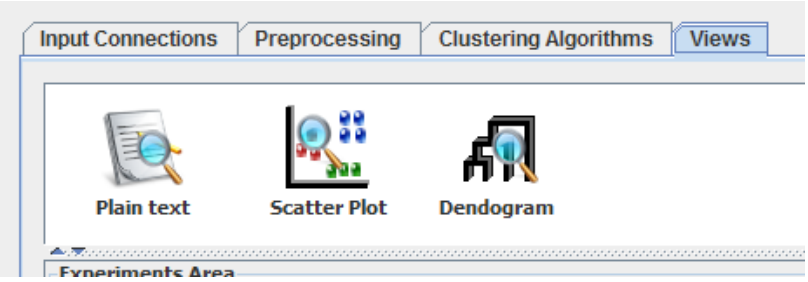
Figura 32: configuración icono DBScan.



1.4 Views o Visualización de Resultados

Permite al usuario ver los resultados de los diferentes algoritmos de clustering aplicados. Se manejan tres tipos de opciones que permiten analizar los resultados desde diferentes puntos de vista. Son fácil manejo y permiten evaluar la calidad de los segmentos formados. Entre ellos se tiene la opcion *Plan text*, que despliega los resultados en formato plano y tablas, la opcion *Scatter Plot* que le presenta al usuario los diferentes tipo de diagras de dispersion y por ultimo que es esclusivo de los algoritmos jerarquicos que la presentacion del diagrama resultado. Ver figura 33.

Figura 33: opciones de vsualización de resultados.

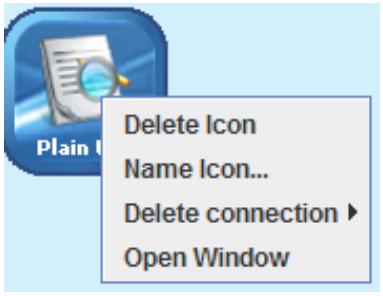


1.4.1 icono Plain text



Permite ver un informe de resultados de la ejecución de un algoritmo de clustering. Sus opciones se presentan en la figura 34.

Figura 34: Opciones icono Plain Text.



Open Window

Permite al usuario desplegar una ventana, que está compuesta por diferentes pestañas que permiten hacer un análisis diferente de los datos. La primera pestaña es *Plain Text* que presenta un informe en donde se tienen aspecto como el error cuadrático, los representantes de clúster, la desviación estándar de los segmentos formados y la distancia euclidiana entre los representantes de clúster. Este informe se puede exportar a archivos con formato .PDF, .RFT y HTML (ver figura 35).

La siguiente pestaña se denomina *Data Assigment*, que permite ver en que segmento fue asignado determinado registro (ver figura 36).

Otra opción es la pestaña *Representatives of Cluster* que presenta los representantes de los segmentos formados.(ver figura 37)

Figura 35: Imagen de la opción del icono Plain Text , pestaña *Plain Text*

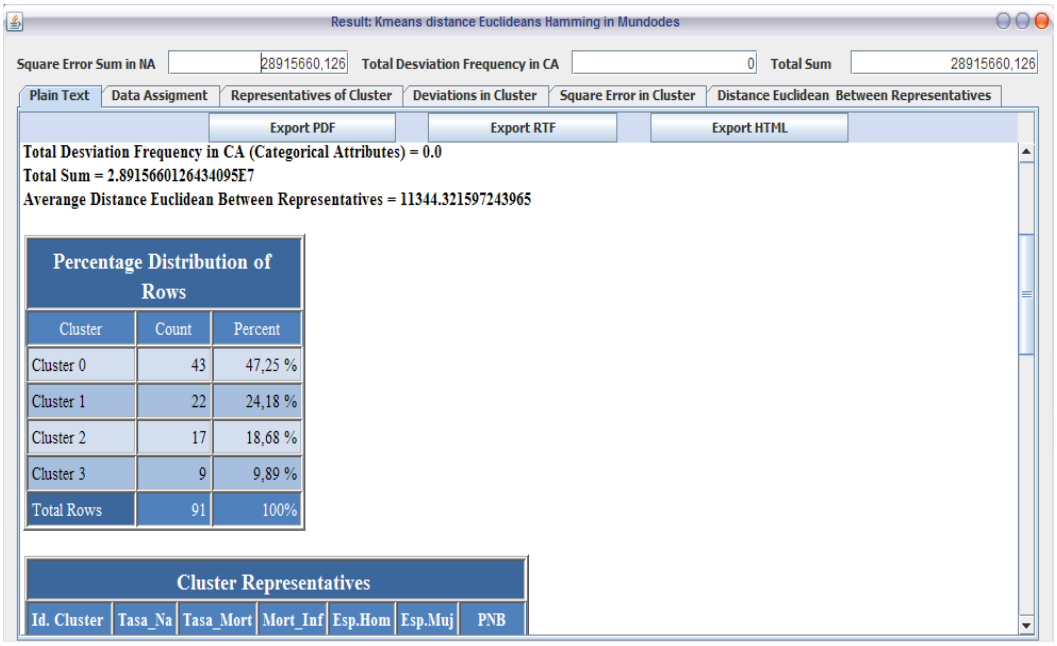
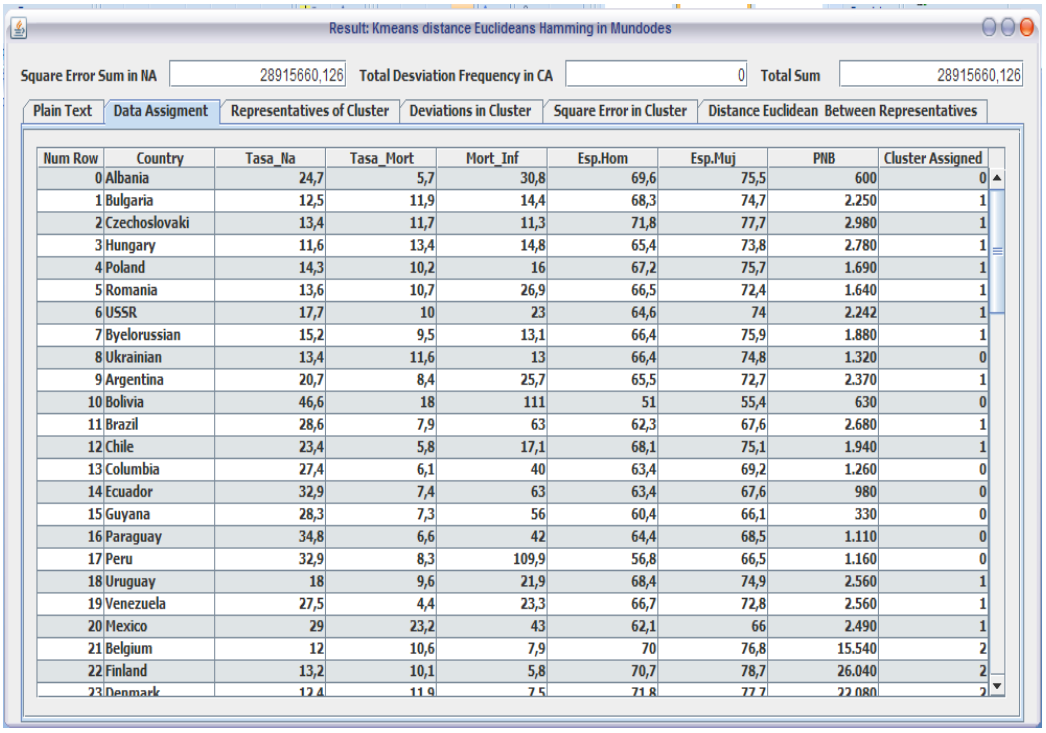
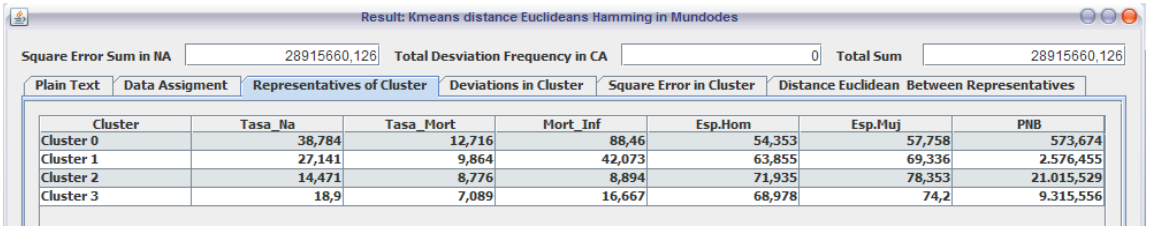


Figura 36: Imagen de la opción del icono Plain Text , pestaña *Data Assigment*



Num Row	Country	Tasa_Na	Tasa_Mort	Mort_Inf	Esp.Hom	Esp.Muj	PNB	Cluster Assigned
0	Albania	24,7	5,7	30,8	69,6	75,5	600	0
1	Bulgaria	12,5	11,9	14,4	68,3	74,7	2.250	1
2	Czechoslovaki	13,4	11,7	11,3	71,8	77,7	2.980	1
3	Hungary	11,6	13,4	14,8	65,4	73,8	2.780	1
4	Poland	14,3	10,2	16	67,2	75,7	1.690	1
5	Romania	13,6	10,7	26,9	66,5	72,4	1.640	1
6	USSR	17,7	10	23	64,6	74	2.242	1
7	Byelorussian	15,2	9,5	13,1	66,4	75,9	1.880	1
8	Ukrainian	13,4	11,6	13	66,4	74,8	1.320	0
9	Argentina	20,7	8,4	25,7	65,5	72,7	2.370	1
10	Bolivia	46,6	18	111	51	55,4	630	0
11	Brazil	28,6	7,9	63	62,3	67,6	2.680	1
12	Chile	23,4	5,8	17,1	68,1	75,1	1.940	1
13	Columbia	27,4	6,1	40	63,4	69,2	1.260	0
14	Ecuador	32,9	7,4	63	63,4	67,6	980	0
15	Guyana	28,3	7,3	56	60,4	66,1	330	0
16	Paraguay	34,8	6,6	42	64,4	68,5	1.110	0
17	Peru	32,9	8,3	109,9	56,8	66,5	1.160	0
18	Uruguay	18	9,6	21,9	68,4	74,9	2.560	1
19	Venezuela	27,5	4,4	23,3	66,7	72,8	2.560	1
20	Mexico	29	23,2	43	62,1	66	2.490	1
21	Belgium	12	10,6	7,9	70	76,8	15.540	2
22	Finland	13,2	10,1	5,8	70,7	78,7	26.040	2
23	Denmark	12,4	11,9	7,5	71,8	77,7	22.080	2

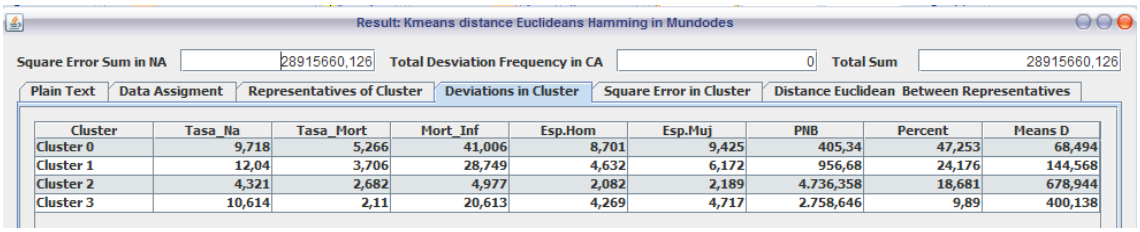
Figura 37: Imagen de la opción del icono Plain Text , pestaña *Representatives of Cluster.*



Cluster	Tasa_Na	Tasa_Mort	Mort_Inf	Esp.Hom	Esp.Muj	PNB
Cluster 0	38,784	12,716	88,46	54,353	57,758	573,674
Cluster 1	27,141	9,864	42,073	63,855	69,336	2.576,455
Cluster 2	14,471	8,776	8,894	71,935	78,353	21.015,529
Cluster 3	18,9	7,089	16,667	68,978	74,2	9.315,556

Otra opción es la pestaña *Deviations in Cluster* que presenta las desviaciones estándar para los atributos numéricos y el error de frecuencia de los atributos categóricos.(ver figura 37)

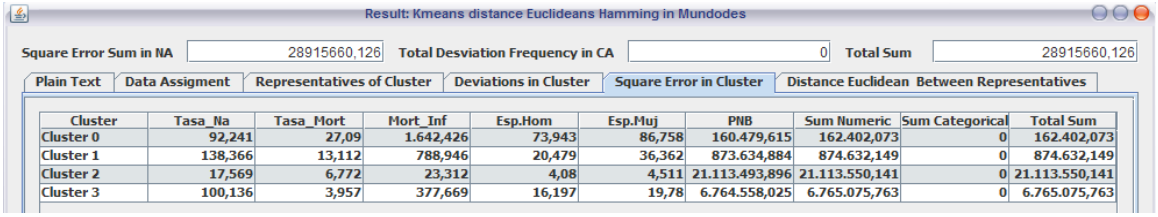
Figura 38: Imagen de la opción del icono Plain Text , pestaña *Deviations in Cluster.*



Cluster	Tasa_Na	Tasa_Mort	Mort_Inf	Esp.Hom	Esp.Muj	PNB	Percent	Means D
Cluster 0	9,718	5,266	41,006	8,701	9,425	405,34	47,253	68,494
Cluster 1	12,04	3,706	28,749	4,632	6,172	956,68	24,176	144,568
Cluster 2	4,321	2,682	4,977	2,082	2,189	4.736,358	18,681	678,944
Cluster 3	10,614	2,11	20,613	4,269	4,717	2.758,646	9,89	400,138

Otra opción es la pestaña *Square Error in Cluster* que presenta los errores cuadráticos promedio de los atributos de tipo numérico y el error de frecuencia de los atributos categóricos.(ver figura 39)

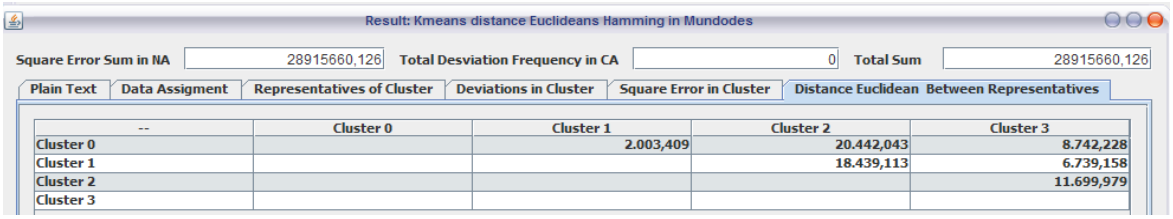
Figura 39: Imagen de la opción del icono Plain Text , pestaña *Square Error in Cluster.*



Cluster	Tasa_Na	Tasa_Mort	Mort_Inf	Esp.Hom	Esp.Muj	PNB	Sum Numeric	Sum Categorical	Total Sum
Cluster 0	92,241	27,09	1.642,426	73,943	86,758	160.479,615	162.402,073	0	162.402,073
Cluster 1	138,366	13,112	788,946	20,479	36,362	873.634,884	874.632,149	0	874.632,149
Cluster 2	17,569	6,772	23,312	4,08	4,511	21.113.493,896	21.113.550,141	0	21.113.550,141
Cluster 3	100,136	3,957	377,669	16,197	19,78	6.764.558,025	6.765.075,763	0	6.765.075,763

Y la ultima pestaña es *Distance Euclidean Between Representatives* que presenta la distancia euclidiana para los numéricos y haming para los categóricos, entre los representantes de grupo. (ver figura 40)

Figura 40: Imagen de la opción del icono Plain Text , pestaña *Distance Euclidean Between Representatives.*



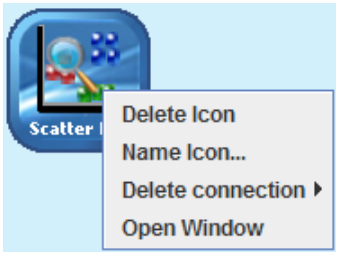
--	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Cluster 0		2.003,409	20.442,043	8.742,228
Cluster 1			18.439,113	6.739,158
Cluster 2				11.699,979
Cluster 3				

1.4.1 Icono Scatter Plot



Permite ver un diagrama de dispersión X vs Y de los atributos, diferenciando la asignación de los segmentos formados. Sus opciones se presentan en la figura 41.

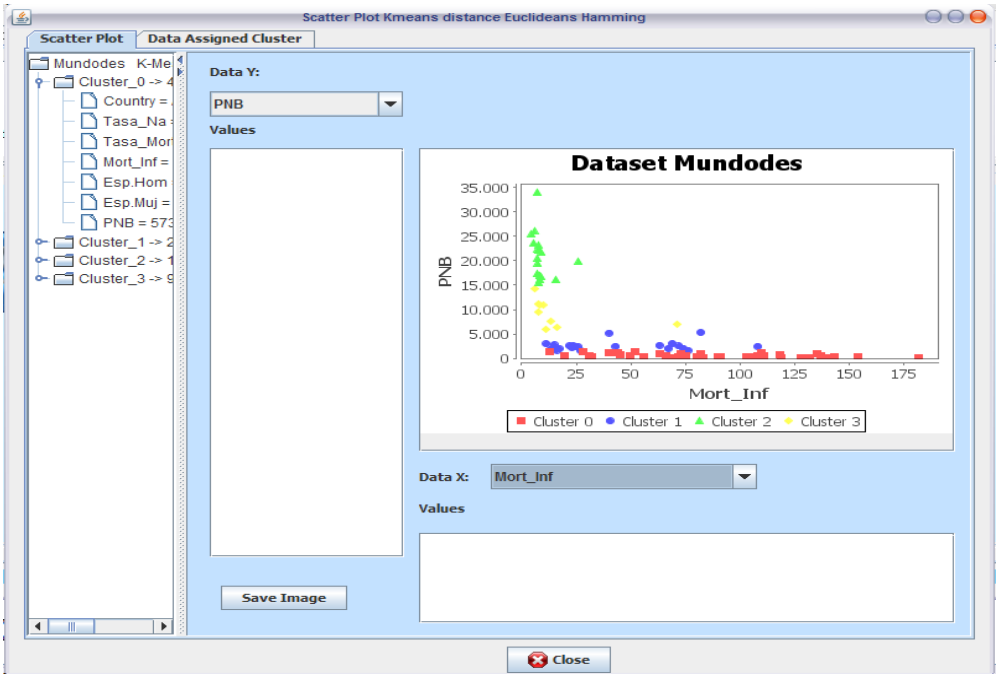
Figura 41: Opciones icono Scatter Plot.



Open Window

Despliega una ventana que le permite interactuar al usuario en la creación de diagramas de dispersión X vs Y, brindándole la posibilidad de guardar las imágenes en formato .JPG.

Figura 42: Visualización de Diagramas de dispersión.

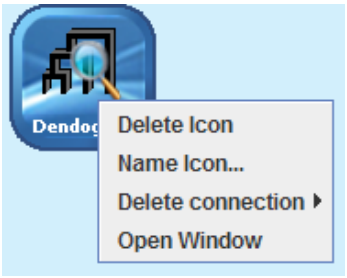


1.4.1 Icono Dendogram



Permite ver el dendograma formado por los algoritmos de clustering jerárquico, Sus opciones se presentan en la figura 43.

Figura 43: Opciones icono Scatter Plot.



Open Window

Permite al usuario analizar la jerarquía generada en los algoritmos de clustering jerárquico diferenciando segmentos formados. Si el usuario al ejecutar los algoritmos jerárquicos coloco un numero de segmentos a formar, estos se presentaran con diferentes colores, además de hacer un efecto de zoom para mejor comprensión del grafico con el Scroll del Mouse, también le da la posibilidad de guardar las imágenes en formato .JPG. Ver figura 44.

Figura 44: Visualización del Dendograma.

