	UNIVERSIDAD DON BOSCO FACULTAD DE INGENIERIA ESCUELA DE COMPUTACION GUIA DE LABORATORIO N° 9
	Nombre de la práctica: Reglas de Asociación Materia: Data WareHouse y Minería de Datos

Reglas de asociación.

Introducción

Dentro de la inteligencia de negocios una de las herramientas más importantes que podemos utilizar, es el uso de técnicas de Minería de Datos, con los cuales cualquier empresa o institución puede obtener resultados importantes para la toma de decisiones.

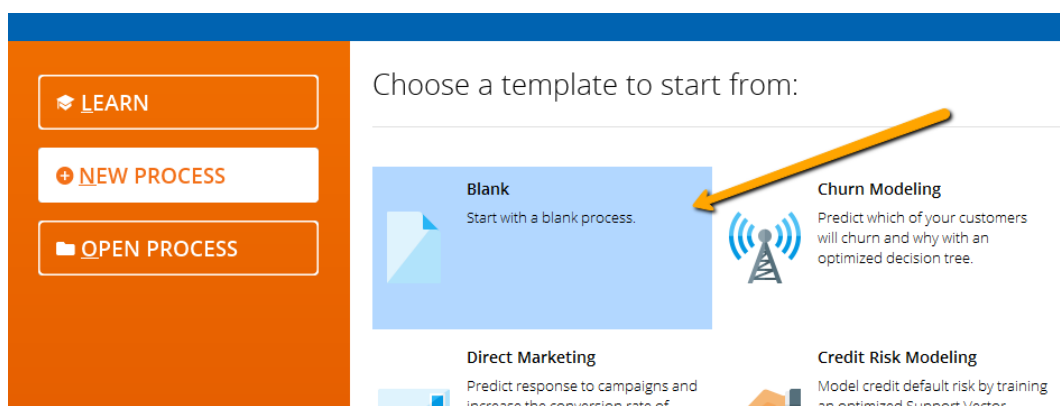
Como este es un tema muy amplio y complejo, lo que veremos haciendo uso de herramienta de analítica de mucho uso en la actualidad como lo es **RapidMiner Studio**, el cual proporciona una versión gratis para análisis de pequeñas empresas.

I. DESARROLLO

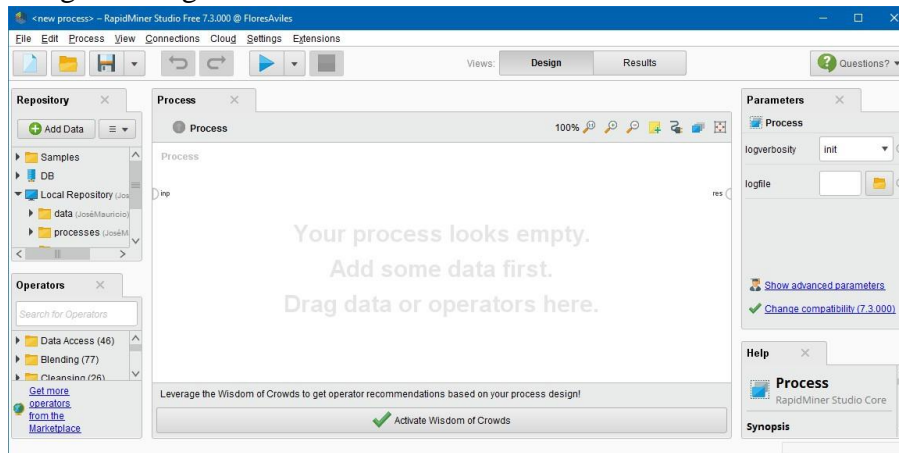
Ejemplo de Aplicación

Realizaremos un ejemplo basado en **“Reglas de Asociación”** para cual iniciaremos con el recurso **datacomestibles** (disponible en el sitio de las guías de laboratorio) y creando una carpeta con el nombre **“reglas”** y guardaremos ahí el archivo que contiene la data el cual es extensión **“csv”**, y el programa **RapidMiner** el cual ya está instalado en su computadora.

Para crear un nuevo lienzo utilizaremos la opción **“Ctrl N”** y seleccionaremos la opción **“Blank”**, como lo muestra la siguiente figura.

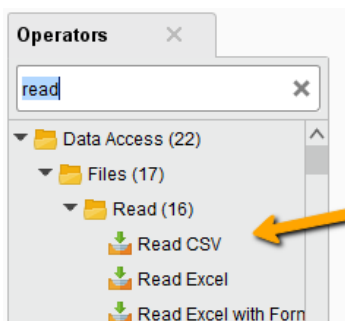


Este proceso nos presentara un nuevo lienzo, para empezar a crear el modelo de datos, como lo muestra la siguiente figura.

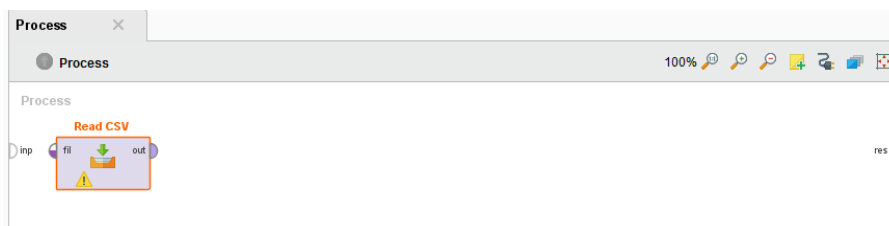


OPERADOR 1 – Read CSV

Ahora vamos seleccionar nuestro primer operador el cual será **“Read CSV”**, para localizar este operador entre los más de 300 operadores, usaremos el buscador de operadores y digitaremos **“read”** y cual nos mostrara todas las coincidencias.

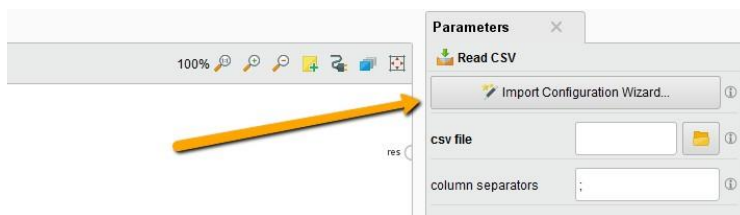


Para ocupar el operador tenemos dos formas de hacerlo, ya sea tomando y arrastrando el operador hacia el lienzo o dándole doble clic.

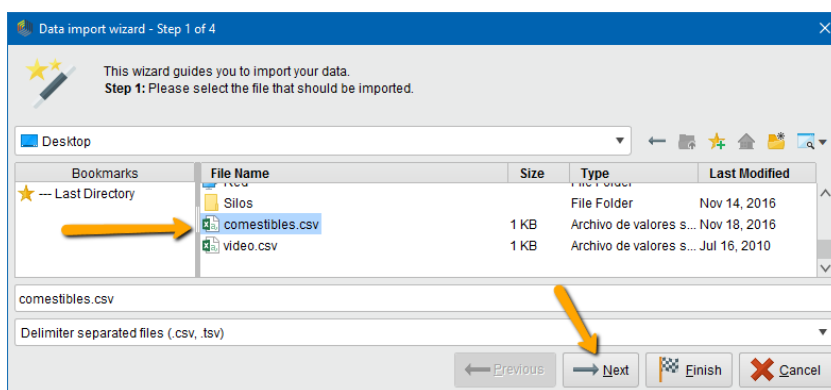


Una vez colocado el operador, lo primeros que haremos es borrar la línea que conecta a nuestro operador con el lado derecho del lienzo, para hacer esto, seleccionamos la línea y presionamos suprimir, ahora vamos a usar los datos de prueba, y ocuparemos la parte derecha

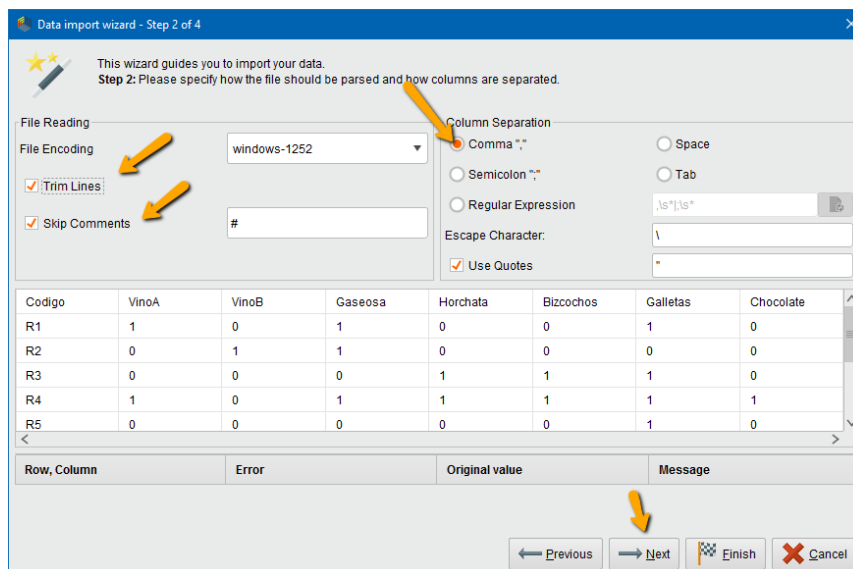
de la aplicación que donde esta los parámetros de cada operador, y seleccionaremos la opción ***“Import Configuration Wizard”*** como lo muestra la siguiente figura.



Recuerde en que carpeta descomprimió el archivos de datos, pues tendrá que llamar al archivo comestibles.csv, como lo muestra la siguiente figura.



Una vez seleccionado el archivo, presionaremos “Next” para continuar con la configuración del modelo, y nos presentara la siguiente pantalla.



Verificamos que la pantalla tenga la configuración correcta, no se debe de cambiar los datos de esta, porque el ejemplo se ha configurado de tal forma, pero en futuros caso puede ser necesario modificar algún parámetro, dependiendo de cómo estos sean obtenidos.

Ahora presionaremos **“Next”** y nos presentara la siguiente pantalla.

Data import wizard - Step 3 of 4

This wizard guides you to import your data.
Step 3: In RapidMiner Studio, each attribute can be annotated. The most important annotation of an attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.

Annotat...	att1	att2	att3	att4	att5	att6	att7	att8
Name	Codigo	VinoA	VinoB	Gaseosa	Horchata	Bizcochos	Galletas	Chocolate
-	R1	1	0	1	0	0	1	0
-	R2	0	1	1	0	0	0	0
-	R3	0	0	0	1	1	1	0
-	R4	1	0	1	1	1	1	1
-	R5	0	0	0	0	0	1	0

Previous Next Finish Cancel

Esta pantalla confirmara la información que se incorporara al modelo, para continuar presiono **“Next”**.

Data import wizard - Step 4 of 4

This wizard guides you to import your data.
Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of your attributes. Furthermore, RapidMiner Studio assigns roles to the attributes, defining what they can be used for by the individual operators. These roles can be also defined here. Finally, you can rename attributes or deselect them entirely.

Reload data Guess value types Date format Enter value...

☒ Preview uses only first 100 rows.

	Codigo	VinoA	VinoB	Gaseosa	Horchata	Bizcochos	Galletas	Chocolate
polyno...	integer	integer	integer	integer	integer	integer	integer	integer
attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute
R1	1	0	1	0	0	1	0	0
R2	0	1	1	0	0	0	0	0
R3	0	0	0	1	1	1	1	0
R4	1	0	1	1	1	1	1	1

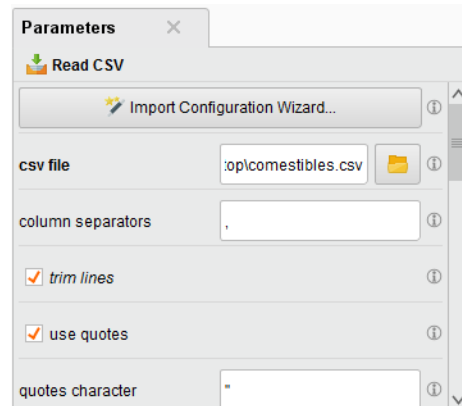
0 errors. ☒ Ignore errors ☐ Show only errors

Row, Column	Error	Original value	Message
-------------	-------	----------------	---------

Previous Next Finish Cancel

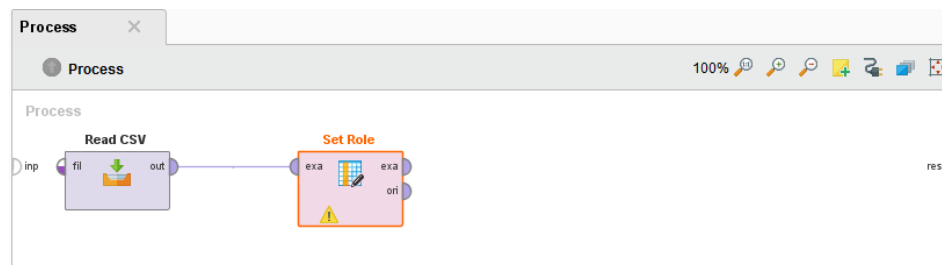
La última pantalla nos permite configurar todavía aún más la data que importamos, para terminar este proceso presionaremos **“Finish”**.

Una vez importado los datos, me aparecerá en el lienzo el operador, solamente con un círculo amarillo y debo de configurar la parte derecha con los datos tal y como aparecen en la siguiente figura.

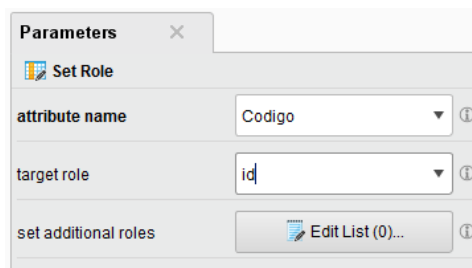


OPERADOR 2 – Set Role

Realizo el mismo proceso con el operador 1, lo busco y al encontrarlo, presiono doble clic o lo arrastro al lienzo, este proceso dejara unidos los dos operadores con una línea, pero también el operador con 2 con el lado derecho del lienzo ***“es importare borrar esta línea, seleccionándola y presionando suprimir”***, al final el modelo se vera de la siguiente manera.



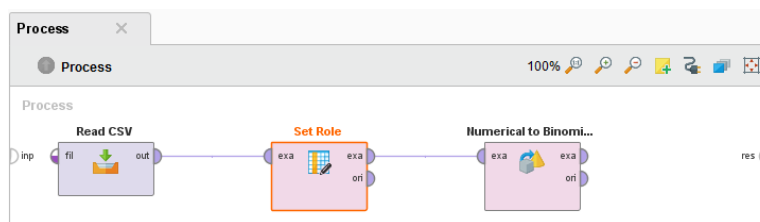
Ahora debemos de configurar los parámetros de del operador, los cuales están a la derecha de la aplicación, los cuales deben de quedar de siguiente forma.



Lo que hacemos es decirle que la columna “*Código*” no debe ser tomada en cuenta pues solamente representa al id de un comprador, y estamos listos para el operador 3.

OPERADOR 3 – Numerical to Binominal

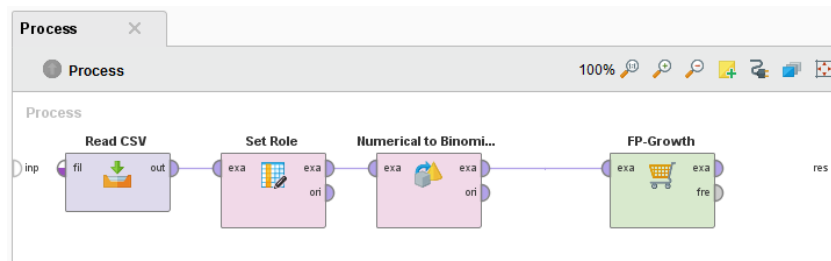
Ahora agregaremos otro operador al modelo, lo buscaremos y lo incorporaremos al lienzo, es importante siempre recordar el borrar la línea que une al operador a la parte derecho del lienzo, la unión entre operador 2 y 3 será automática, es importante ver que el operador 2 tiene 2 salidas, el lienzo nos quedara de la siguiente forma.



En la parte de los parámetros no se cambiara ninguna selección por lo cual debe de quedar de la siguiente manera.

OPERADOR 4 – FP GROWTH

El siguiente operador es el “*FP GROWTH*”, para agregarlo hacemos como los anteriores los buscamos y lo incorporamos al lienzo, y como en los otros casos borramos la línea del operador con la parte derecha del lienzo y nos queda como la siguiente figura.



Ahora configuramos los parámetros del operador, y estos datos deben de quedar con la siguiente figura.

The screenshot shows the 'Parameters' window for the 'FP-Growth' operator. The parameters are as follows:

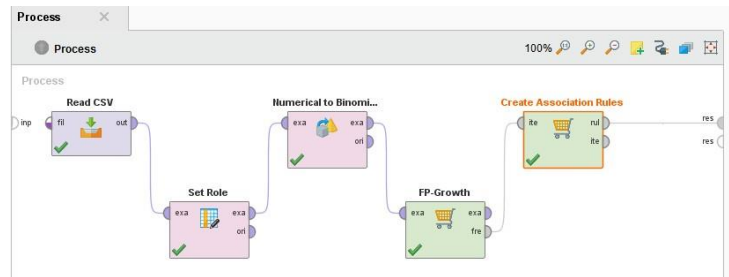
Parameter	Value
find min number of itemsets	<input checked="" type="checkbox"/>
min number of itemsets	100
max number of retries	15
positive value	
min support	0.95
max items	-1
must contain	

La línea de unión se realizara en forma automática, ahora solo nos falta incorporar el ultimo operador.

OPERADOR 5 – Create Association Rules

De la misma forma en que hemos agregado los demás operadores, agregaremos el último, que es “*Create Association Rules*” el cual nos servirá para configurar que reglas y que valores necesitamos que aplique el modelo.

En este operador SI dejaros la línea que conecta el operador con la parte derecha del lienzo, para que pueda generar resultados según nuestras condiciones. El lienzo completo se vería de esta forma.



Y debemos también configurar los parámetros, y realizaremos una prueba con la siguiente configuración

Parameter	Value
criterion	confidence
min confidence	0.5
gain theta	2.0
laplace k	1.0

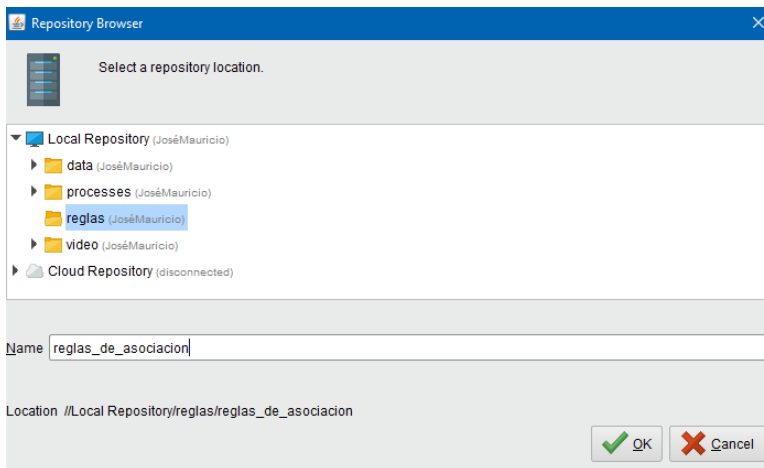
EJECUTAR EL MODELO

Para que se realice todo el proceso, ejecutamos el modelo presionando el icono de Play azul, el cual nos servirá para tener los datos disponibles, también ocuparemos los iconos de libreta de apuntes que servirá para regresar al modelo, y el icono de presentación que servirá para ir al modelo.

No.	Premises	Conclusion
10	Gaseosa	Bizcochos

La primera vez que lo corramos nos pedirá que lo guardemos, a lo cual accederemos, y nos pedirá en que repositorio lo alojaremos, si el repositorio no está a la vista usemos el que tenemos por default.

El nombre que usaremos para este ejemplo es: reglas_asociacion_ejemplo, pero puede usar otro nombre, la información debe estar colocada con la pantalla siguiente.



Result History

AssociationRules (Create Association Rules)

Show rules matching: all of these conclusions:

Min. Criterion: confidence

Min. Criterion Value: 0.3

No.	Premises	Conclusion	Support	Conf.
10	Gaseosa	Bizcochos	0.400	0.66
11	Bizcochos	Gaseosa	0.400	0.66
12	Bizcochos	Horchata	0.400	0.66
13	VinoA	Gaseosa	0.300	0.75
14	VinoA	Galletas	0.300	0.75
15	Horchata	Galletas	0.300	0.75
16	Chocolate	Galletas	0.300	0.75
17	Chocolate	Bizcochos	0.300	0.75
18	VinoA	Chocolate	0.300	0.75
19	Chocolate	VinoA	0.300	0.75
20	Horchata	Galletas, Bizcochos	0.300	0.75

Es muy difícil en una guía mostrar todos los resultados, entonces debemos de experimentar utilizando las opciones de la parte izquierda, seleccionando valores, cambiando criterios, o cambiando el valor mínimo del criterio, esto nos puede generar muchas opciones y resultados, además debemos de analizar los datos de la lista.

II. ANÁLISIS DE RESULTADOS

1. Aplique el mismo proceso pero ahora con el recurso `ferreteria.csv` y cree un informe de resultados, ¿qué datos interesantes nos arroja este archivo?

III. TAREA COMPLEMENTARIA

1. Explore el sitio <https://www.mockaroo.com/> , el cual provee datos en formato csv para realizar pruebas. Investigue qué tipo de datos se encuentran alojados y cómo pueden ser utilizados. Descargue un csv a partir de la información que nos provee y aplíquelo reglas de asociación. Presente y analice la información que nos muestra utilizando RapidMiner