



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA

Laboratorio de:

ANALÍTICA DE DATOS – BIG DATA

Práctica No.: 13

Tema: Rapidminer

Objetivos:

- Realizar los comandos del tutorial
- Responder las preguntas propuestas en este laboratorio.

Marco teórico:

RapidMiner [1]

Es un programa informático para el análisis y minería de datos. Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico. Se usa en investigación, educación, capacitación, creación rápida de prototipos y en aplicaciones empresariales. En una encuesta realizada por KDnuggets, un periódico de minería de datos, RapidMiner ocupó el segundo lugar en herramientas de analítica y de minería de datos utilizadas para proyectos reales en 2009¹ y fue el primero en 2010. La versión inicial fue desarrollada por el departamento de inteligencia artificial de la Universidad de Dortmund en 2001. Se distribuye bajo licencia AGPL y está hospedado en SourceForge desde el 2004.

RapidMiner proporciona más de 500 operadores orientados al análisis de datos, incluyendo los necesarios para realizar operaciones de entrada y salida, preprocesamiento de datos y visualización. También permite utilizar los algoritmos incluidos en Weka.

Desarrollo de la práctica:

Dentro del DataBase de examples Training insertar la Transacciones y productos



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA

The screenshot shows the RapidMiner Studio interface. The 'Process' view displays a workflow with two operators: 'Retrieve Products' and 'Retrieve Transactions'. The 'Repository' view on the left lists various data sources, including 'Ripley-Set (v1)', 'Sonar (v1)', 'Titanic (v1)', 'Titanic Training (v1)', 'Titanic Unlabeled (v1)', 'Transactions (v1)', and 'Weighting (v1)'. The 'Parameters' view on the right shows the settings for the 'Process' operator, including 'logverbosity' set to 'init' and 'logfile'. The 'Help' view on the right shows the documentation for the 'Process' operator, which is described as 'The root operator which is the outer most operator of every process.'

Posterior insertamos el operador de Join para unir las dos BD.

The screenshot shows the RapidMiner Studio interface with the 'Join' operator added to the workflow. The 'Process' view displays the 'Join' operator connected to the outputs of 'Retrieve Products' and 'Retrieve Transactions'. The 'Parameters' view on the right shows the settings for the 'Join' operator, including 'join type' set to 'inner' and 'key attributes' set to 'id'. The 'Help' view on the right shows the documentation for the 'Join' operator, which is described as 'This Operator joins two ExampleSets using one or more Attributes of the input ExampleSets as key attributes.'

Editamos los parámetros de Join



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA

RapidMiner Studio Trial 9.3.000 @ DESKTOP-PQ7B8TF

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

Repository

- Import Data
- Ripley-Set (v1)
- Sonar (v1)
- Titanic (v1)
- Titanic Training (v1)
- Titanic Unlabeled (v1)
- Transactions (v1)
- Weighting (v1)
- processes
- Templates

Operators

- Join
- Blending (7)
- Table (7)
- Joins (7)
- Append
- Join
- Set Minus
- Intersect

We found "Data Search for Data Mining" in the Marketplace. [Show me!](#)

Edit Parameter List: key attributes

Edit Parameter List: key attributes
The attributes which shall be used for join. Attributes which shall be matched must be of the same type.

left key attributes	right key attributes
Product ID	Product ID

Add Entry Remove Entry Apply Cancel

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds

Parameters

Join

join type: inner

use id attribute as key

key attributes: Edit List (0)

Show advanced parameters

Help

Join

Concurrency

Tags: Combine, Merge, Match, Left, Right, Outer, Inner, Joins

Synopsis

This Operator joins two ExampleSets using one or more Attributes of the input ExampleSets as key attributes.

Insertamos el aggregate y sus parámetros.

RapidMiner Studio Trial 9.3.000 @ DESKTOP-PQ7B8TF

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

Repository

- Import Data
- Training Resources (connected)
- Samples
- Community Samples (connected)
- DB (Legacy)
- Local Repository (Ale)

Operators

- Data Access (53)
- Blending (79)
- Cleansing (25)
- Modeling (156)
- Scoring (12)
- Validation (29)
- Utility (86)

Get more operators from the Marketplace

Process

Process

100%

Retrieve Products

Retrieve Transactions

Join

Aggregate

Parameters

Process

logverbosity: int

logfile

Show advanced parameters

Change compatibility (9.3.000)

Help

Process

RapidMiner Studio Core

Synopsis

The root operator which is the outer most operator of every process.

Description



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA

RapidMiner Studio Trial 9.3.000 @ root-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

Repository

- Import Data
- Purchases (v1)
- Ripley-Set (v1)
- Sonar (v1)
- Titanic (v1)
- Titanic Training (v1)
- Titanic Unlabeled (v1)
- Transactions (v1)
- Weighting (v1)
- processes
- Templates
- Time Series
- Tutorials

Operators

Agg

- Blending (4)
- Attributes (1)
- Generation (1)
- Generate Aggregation
- Table (3)
- Grouping (1)
- Aggregate
- Rotation (2)
- Pivot
- De-Pivot

No results were found.

Process

Process

100%

Select Attributes: group by attributes

Select Attributes: group by attributes
Performs a grouping by the values of the attributes by the selected attributes.

Attributes

- Search
- Amount
- Customer ID
- Price
- Product Category
- Product ID
- Product Name

Selected Attributes

- Search

Apply Cancel

Recommended Operators

- Select Attributes 51%
- Filter Examples 40%
- Multiply 35%
- Set Role 34%

Parameters

Aggregate

- use default aggregation
- aggregation attributes Edit List (0)...
- group by attributes Select Attributes...

Show advanced parameters

Change compatibility (9.3.000)

Help

Aggregate

RapidMiner Studio Core

Tags: Groupby, Group by, Grouping, Sum, Count, Min, Max, Average, Avg, Mean, Pivot, Cross-table, Crosstable, Distinct, Percentile

Synopsis

This operator performs the aggregation functions known from SQL. This operator provides a lot of functionalities in the same format as provided by the SQL aggregation functions. SQL aggregation functions

RapidMiner Studio Trial 9.3.000 @ root-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

Repository

- Import Data
- Purchases (v1)
- Ripley-Set (v1)
- Sonar (v1)
- Titanic (v1)
- Titanic Training (v1)
- Titanic Unlabeled (v1)
- Transactions (v1)
- Weighting (v1)
- processes
- Templates
- Time Series
- Tutorials

Operators

Agg

- Blending (4)
- Attributes (1)
- Generation (1)
- Generate Aggregation
- Table (3)
- Grouping (1)
- Aggregate
- Rotation (2)
- Pivot
- De-Pivot

No results were found.

Process

Process

100%

Select Attributes: group by attributes

Select Attributes: group by attributes
Performs a grouping by the values of the attributes by the selected attributes.

Attributes

- Search
- Amount
- Customer ID
- Price
- Product Category
- Product ID
- Product Name

Selected Attributes

- Search
- Product ID

Apply Cancel

Recommended Operators

- Select Attributes 51%
- Filter Examples 40%
- Multiply 35%
- Set Role 34%

Parameters

Aggregate

- use default aggregation
- aggregation attributes Edit List (0)...
- group by attributes Select Attributes...

Show advanced parameters

Change compatibility (9.3.000)

Help

Aggregate

RapidMiner Studio Core

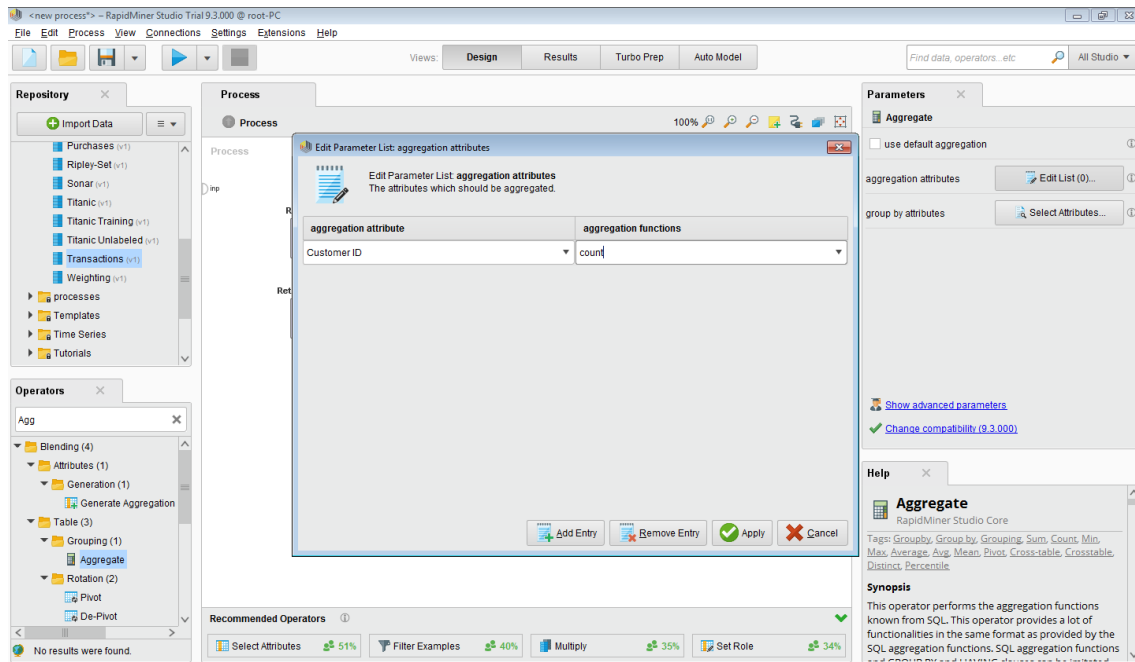
Tags: Groupby, Group by, Grouping, Sum, Count, Min, Max, Average, Avg, Mean, Pivot, Cross-table, Crosstable, Distinct, Percentile

Synopsis

This operator performs the aggregation functions known from SQL. This operator provides a lot of functionalities in the same format as provided by the SQL aggregation functions. SQL aggregation functions



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA



Respuestas

1. ¿Qué producto se vendió con mayor frecuencia?

Row No.	maximum(count(Customer ID))	mode(mode(Product Na...))
1	22	Stim pack

ExampleSet (1 example, 0 special attributes, 2 regular attributes)

2. ¿Qué producto se vendió solo 5 veces?



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA

ExampleSet (Filter Examples)

Open in: Turbo Prep | Auto Model

Filter (1 / 1 examples): all

Row No.	Product ID	count(Custo...	mode(Produ...
1	129	5	Bacta

ExampleSet (1 example, 0 special attributes, 3 regular attributes)

3. ¿Puede averiguar en la pestaña Estadísticas cuál fue el número promedio de transacciones?

ExampleSet (Aggregate)

Filter (3 / 3 attributes): Search for Attributes

Name	Type	Missing	Statistics
Product ID	Integer	0	Min: 1, Max: 178, Average: 89.500
count(Customer ID)	Integer	0	Min: 4, Max: 22, Average: 13.079, Deviation: 3.623
mode(Product Name)	Polynomial	0	Least: Zydrate (1), Most: Stim pack (2), Values: Stim pack (2), ALZ-112 (1), ... [175 more]

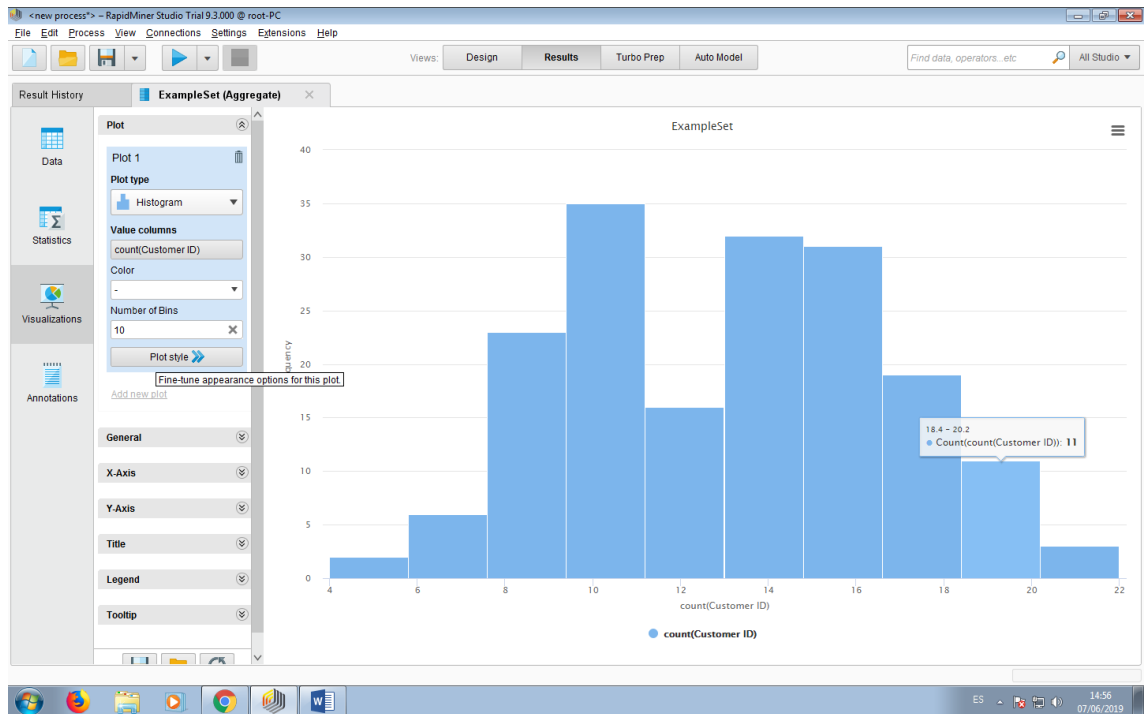
Showing attributes 1 - 3

Examples: 178 Special Attributes: 0 Regular Attributes: 3

4. ¿Puedes ver también la distribución visual de los valores en esta pestaña?



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA



5. La función de conteo cuenta el número de transacciones para cada producto, pero cada producto también se puede comprar varias veces en cada transacción.

ExampleSet (Aggregate)

Open in: Turbo Prep, Auto Model

Filter (178 / 178 examples): all

Row No.	Product ID	count(Custo...	mode(Produ...	sum(Amount)
1	1	19	Repressitol	55
2	2	8	Ritalout	30
3	3	9	Comanapracil	23
4	4	14	Serum 114	37
5	5	9	Hypnocil	28
6	6	13	Substance D	31
7	7	13	Hydronium Br...	43
8	8	12	Azoth	39
9	9	12	Digitalin	26
10	10	7	Stim pack	23
11	11	11	Damascus R...	39
12	12	15	Nepenthe	38
13	13	10	Blaccine	34
14	14	12	Teamocil	40
15	15	11	Metazine	33
16	16	11	Chamaila ext...	31
17	17	19	Doloxan	43
18	18	15	Morpha	55
19	19	17	Serisone	44

ExampleSet (178 examples, 0 special attributes, 4 regular attributes)

6. ¿Puedes cambiar los parámetros de Agregado para que se calcule la suma total de cada producto?



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA

RapidMiner Studio Trial 9.3.000 @ DESKTOP-PQ788TF

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Result History ExampleSet (Aggregate)

Open in Turbo Prep Auto Model

Filter (178 / 178 examples): all

Row No.	Product ID	sum(Amount)	mode(Produ...
1	1	55	Repressitol
2	2	30	Ritalout
3	3	23	Comanapracil
4	4	37	Serum 114
5	5	28	Hypnocil
6	6	31	Substance D
7	7	43	Hydronium Br...
8	8	39	Azoth
9	9	26	Digitalin
10	10	23	Slim pack
11	11	39	Damascus R...
12	12	38	Nepenthe
13	13	34	Biaccine
14	14	40	Teamocil
15	15	33	Metazine

ExampleSet (178 examples, 0 special attributes, 3 regular attributes)

Repository

- Import Data
- Golf (v1)
- Golf-TestSet (v1)
- Iris (v1)
- Labor-Negotiations (v1)
- Market-Data (v1)
- Polynomial (v1)
- Products (v1)
- Purchases (v1)
- Ripley-Set (v1)
- Sonar (v1)
- Titanic (v1)
- Titanic Training (v1)
- Titanic Unlabeled (v1)
- Transactions (v1)
- Weighting (v1)
- processes
- Templates
- Time Series
- Tutorials
- Community Samples (connected)
- DB (Legacy)
- Local Repository (A1e)

7. ¿Qué productos se han vendido más de 65 veces?

RapidMiner Studio Trial 9.3.000 @ DESKTOP-PQ788TF

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Repository

- Import Data
- Ripley-Set (v1)
- Sonar (v1)
- Titanic (v1)
- Titanic Training (v1)
- Titanic Unlabeled (v1)
- Transactions (v1)
- Weighting (v1)
- processes
- Templates

Operators

- filter
- Examples (2)
 - Filter (2)
 - Filter Examples
 - Filter Example Range
 - Cleansing (1)
 - Duplicates (1)
 - Remove Duplicates
- We found "Text Processing" and "Holt-Winters Filtering" in the Marketplace. [Show me!](#)

Click to select, drag to move.

Create Filters: filters

Create Filters: filters
Defines the list of filters to apply.

count(Customer ID) ≥ 65

Match all Match any Preselect comparators

Add Entry OK Cancel

Activate Wisdom of Crowds

Filter Examples

RapidMiner Studio Core

Select, Keep, Remove, Drop, Delete, Rows, Instances, Lines, Observations, Filter, Filter

This Operator selects which Examples of an ExampleSet are kept and which Examples are



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA

The screenshot shows the RapidMiner Studio interface. The 'ExampleSet (Filter Examples)' window is open, displaying a table with 0 examples. The 'Repository' panel on the right shows a list of data sources, including 'Golf (v1)', 'Golf-Testset (v1)', 'Iris (v1)', 'Labor-Negotiations (v1)', 'Market-Data (v1)', 'Polynomial (v1)', 'Products (v1)', 'Purchases (v1)', 'Ripley-Set (v1)', 'Sonar (v1)', 'Titanic (v1)', 'Titanic Training (v1)', 'Titanic Unlabeled (v1)', 'Transactions (v1)', and 'Weighting (v1)'. The 'Local Repository (Ale)' is selected.

Mayor a 19

The screenshot shows the RapidMiner Studio interface with the 'ExampleSet (Filter Examples)' window displaying 14 examples. The 'Repository' panel on the right is the same as in the previous screenshot.

Row No.	Product ID	count(Custo...	model(Produ...
1	1	19	Repressitol
2	17	19	Doloxan
3	34	20	Zombrex
4	41	19	Senzu Bean
5	42	21	Turbolax
6	67	19	Formula 977
7	79	19	Quinium
8	81	19	E-Z Doze It SL...
9	82	20	Qualex
10	103	20	Safsprin
11	138	22	Althsat
12	155	19	Flintstones C...
13	156	20	Jammilin
14	175	21	Phalanx

Análisis de resultados:

Durante este laboratorio se analizó los conectores de filtro, join y agregación lo cual significa que podemos hacer uso de varias tablas de datos y unirlas, además es importante poner los correctos atributos para los resultados deseados. En el ultimo ejercicio no existe mayores a 65 sin embargo se realizo con un filtro de mayor de 19 y da como resultado que un producto máximo se ha vendido 22 veces.

Conclusiones y recomendaciones:



**ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN
ANALÍTICA DE DATOS
MACAS CEVALLOS ALEXANDRA**

- El uso de herramienta facilita la elaboración de entrenamiento de datos, de una o más tablas.
- Se recomienda usar los demás conectores de RapidMiner
- Se realizó los ejercicios por que existen varias maneras de la visualización de los datos.
- Se observó que también se puede visualizar los resultados de manera gráfica
- Se recomienda leer la documentación de esta herramienta para un mejor entendimiento.

Bibliografía

- [1] «Wikipedia,» [En línea]. Available: <https://es.wikipedia.org/wiki/RapidMiner>. [Último acceso: 06 junio 2019].