



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

---

**Laboratorio de:** ANALÍTICA DE DATOS – BIG DATA

**Práctica No.:** 5

**Tema:** Rapid Miner

**Estudiante:** Danny Sebastián Díaz Padilla

**Objetivos:**

- Instalar la herramienta y familiarizarse con el entorno.
- Realizar una práctica sobre minería de datos en un tema determinado.
- Utilizar el API de Twitter para extraer información de tweets.

**Marco teórico:**

**Minería de datos**

Se puede entender como la tecnología y software utilizado para encontrar patrones de comportamiento dentro de la base de datos. La base fundamental de esto es que esos patrones ayuden a la toma de decisiones. Por ejemplo, podría ayudar a empresas, a conocer los patrones de comportamiento de sus clientes. De manera que le facilitaría el establecimiento de estrategias para incrementar las ventas o reducir costes. [1]

**Rapid Miner**

RapidMiner es una plataforma de análisis que permite acelerar la creación, entrega y mantenimiento de analíticas predictivas de alto valor.

Rapid Miner es una herramienta de Business Intelligence (Inteligencia de Negocios) que proporciona un entorno integrado de aprendizaje automático para la Minería de Datos. Se utiliza para aplicaciones de negocios, para la investigación, educación, formación, creación de prototipos y desarrollo de aplicaciones. [2]

Además, es compatible con todos los pasos del proceso de minería de datos. Esta herramienta incluye:

- Carga y transformación de datos (extracción, transformación, carga ETL)
- Pre-procesamiento de datos y visualización
- Análisis predictivo y modelos estadísticos
- Evaluación y despliegue



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

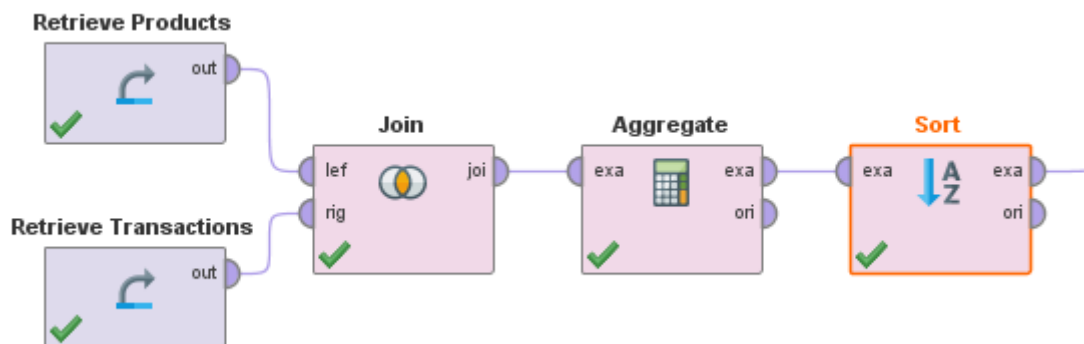
Desarrollo de la práctica:

1. Ejercicios usando Rapid Miner

- ¿Qué producto se vendió con mayor frecuencia?

El **Athsat** con 22 clientes que lo eligieron.

Es necesario realizar una juntura entre la tabla de Productos y la de transacciones para luego realizar un conteo de los clientes que compraron un cierto producto.



Open in Turbo Prep

Auto Model

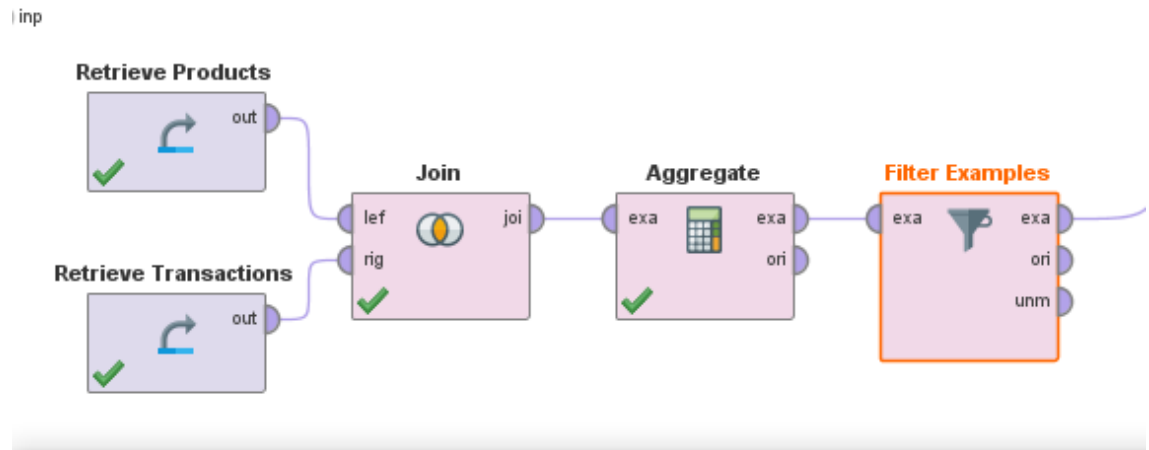
Row No.	Product ID	count(Custo...	mode(Produ...
1	138	22	Athsat
2	42	21	Turbolax
3	175	21	Phalanx
4	34	20	Zombrex
5	82	20	Qualex



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

- ¿Qué producto se vendió solo 5 veces?

El Bacta fue comprado 5 veces por diferentes clientes



Create Filters: filters



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

count(Customer ID)

=

5

Con “Filter” hacemos un “matching” directamente con el que cumple el requisito

Row No.	Product ID	count(Custo...	mode(Produ...	sum(Amount)
1	129	5	Bacta	13

Pero si la pregunta es acerca de cantidades entonces analizamos con amount=5, como resultado ningún producto cumple con la condición

Row No.	Product ID	count(Custo...	mode(Produ...	sum(Amount)



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

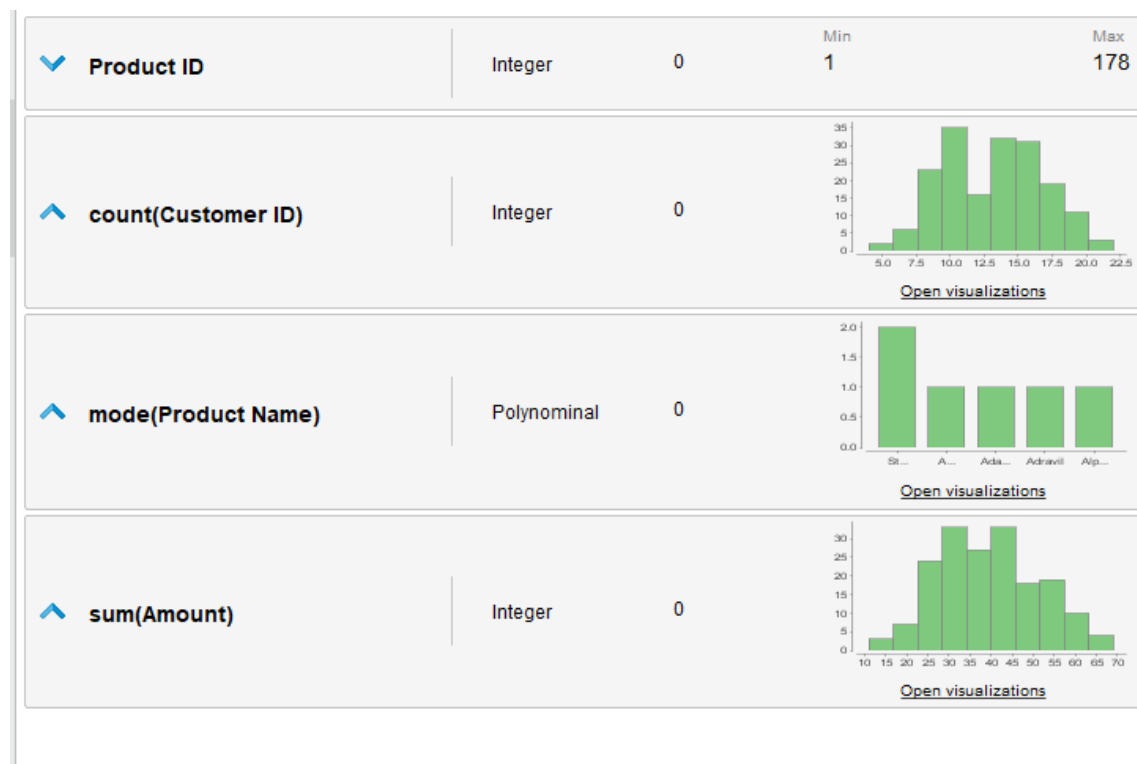
- ¿Puede averiguar en la pestaña Estadísticas cuál fue el número promedio de transacciones?

13.079 transacciones, en este caso cada fila tiene el conteo de transacciones por producto.

✓ Product ID	Integer	0	Min 1	Max 178	Average 89.500
✓ count(Customer ID)	Integer	0	Min 4	Max 22	Average 13.079
✓ mode(Product Name)	Polynomial	0	Least Zydrate (1)	Most Stim pack (2)	Values Stim pack

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

La distribución para el número de clientes y la suma de todos los productos intenta formar una campana de Gauss. Para la moda la visualización muestra algo de uniformidad a partir del segundo valor.





ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

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

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

Para esto se debe agregar la suma de la cantidad por cada producto.

aggregation attribute	aggregation functions
Customer ID ▼	count ▼
Product Name ▼	mode ▼
Amount ▼	sum ▼

Y de esa forma se obtiene la suma total.

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



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

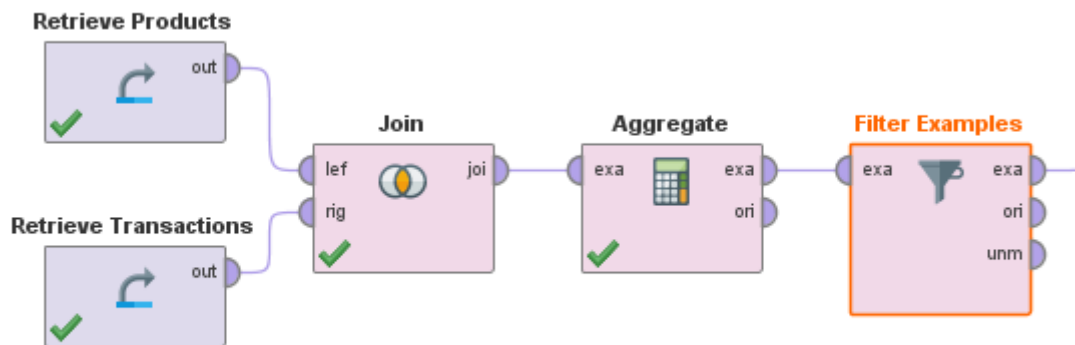
- ¿Qué productos se han vendido más de 65 veces?

Turbolax y Qualex. Filtramos la suma de los productos que son mayores a 65.

Create Filters: filters

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

sum(Amount) > 65



Row No.	Product ID	count(Custo...	mode(Produ...	sum(Amount)
1	42	21	Turbolax	69
2	82	20	Qualex	66



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

## 2. Cosecha de información en Twitter

Primero es necesario tener una cuenta activa de twitter, luego creamos una conexión:

Create a new connection

Connections are now objects that are stored in the repository. You can enter more details to your connection in the next steps.

Connection Type: Twitter

Repository: Local Repository

Connection Name: TwitterEjercicio

Create Cancel

Solicitamos un token y luego de dar todos los permisos nos darán un código

Edit connection - TwitterEjercicio

Info Setup Sources

Access Token

Set injected parameters

Test connection

Leverage the Wisdom of the Crowd

by an external source.

Save Cancel

Syn This

OAuth

Authenticate RapidMiner via OAuth.

1 Open URL

Request access token

Show URL instead

2 Copy Code

Complete Cancel



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

## Authorize RapidMiner SocialMedia Extension to access your account?

☒ Remember me · [Forgot password?](#)

OAuth

Authenticate RapidMiner via OAuth.

1 Open URL

[Show URL instead](#)

2 Copy Code

Una vez aparezca el 'visto' de color verde, significa que la conexión se hizo correctamente.



Para realizar una prueba de la conexión, buscamos los hashtags que contengan la cadena "ParoNacionalEc".

Parameters

☒ Search Twitter

query

To-User-Id	Language	Source	Text	Geo-Locatio...	Geo-Locati...
-1	es	<a href="http://...	#ParoNacionalEc   A alguien le estorbaban documentos de la C...	?	?
-1	es	<a href="http://...	. @virgiliohermand dice que no se irá de Ecuador y que enfren...	?	?
-1	es	<a href="http://...	Por todas las características registradas durante el #ParoNacio...	?	?
-1	es	<a href="http://...	RT @ismaelUlloa11: En el #ParoNacionalEc me llamó la atenci...	?	?
-1	es	<a href="http://...	RT @ecuainm_oficial: ►#ATENCIÓN @CIDH se pronuncia al c...	?	?
-1	es	<a href="http://...	RT @ecuainm_oficial: ►#ATENCIÓN @CIDH se pronuncia al c...	?	?
-1	es	<a href="http://...	RT @eluniversocom: Por todas las características registradas ...	?	?
-1	es	<a href="http://...	RT @GermanyaBravo: Por todas las características registradas...	?	?
-1	es	<a href="http://...	RT @FranYcaza: El Consorcio Jurídico en Defensa de los Afect...	?	?
-1	es	<a href="http://...	RT @FranYcaza: El Consorcio Jurídico en Defensa de los Afect...	?	?
-1	es	<a href="http://...	RT @kolectiVOZ: #ATENCIÓN   #VirgilioHernández: "Pedimos q...	?	?
-1	es	<a href="http://...	RT @ccemanabi: Ya recortaron 400 millones en educación, 20...	?	?
-1	es	<a href="http://...	RT @ecuainm_oficial: ►#ATENCIÓN @CIDH se pronuncia al c...	?	?
-1	es	<a href="http://...	RT @EstebanVillama9: @yakuperezg Le exhortamos a usted s...	?	?





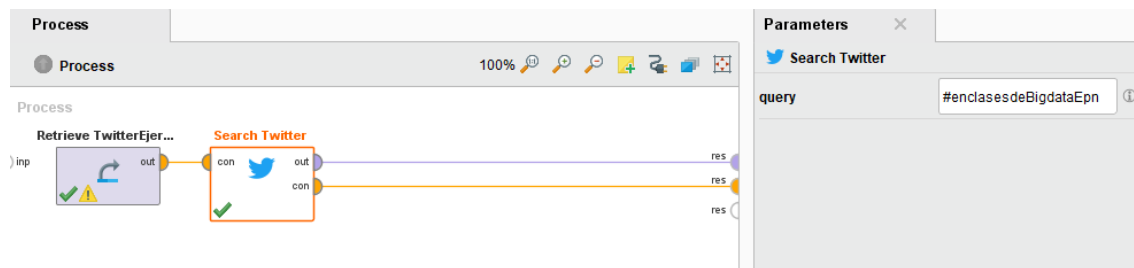
ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

a. Ejercicios

- Desde su cuenta de Twitter crear tweets con los hashtags #enclasesdeBigdataEpn y mencionar a la profesora @cindicita1 y entre sus amigos.



- Buscar cada Hashtag en rapidminer.



En el círculo se encuentra el Twitt que publiqué, demuestra que el análisis es reciente.

Result History						
Connection (Retrieve TwitterEjercicio) ExampleSet (Search Twitter)						
Open in Turbo Prep Auto Model						
Filter (15 / 15 examples): all						
Language	Source	Text	Geo-Locatio...	Geo-Locatio...	Retweet-Co...	
fr	<a href="http...	#enClasesDeBigdataEpn cosecha de tweets con	?	?	0	
es	<a href="http...	Aquí hackeando a la NASA con html con la guía de #encla...	?	?	0	
es	<a href="http...	RT @Henry32299516: #enclasesdeBigdataEpn de la prof...	?	?	1	
es	<a href="http...	ya manden a las casas #enclasesdeBigdataEpn @JonaS...	?	?	1	
es	<a href="http...	#enclasesdeBigdataEpn de la profe @cindicita1 con el Ke...	?	?	1	
und	<a href="http...	:3	?	?	0	
und	<a href="http...	#enclasesdeBigdataEpn	?	?	0	
es	<a href="http...	RT @PedroCuasqui: #enclasesdeBigdataEpn con @alex...	?	?	1	
es	<a href="http...	#enclasesdeBigdataEpn con mi compañero @PedroCua...	?	?	1	
und	<a href="http...	#enclasesdeBigdataEpn @cindicita1	?	?	0	
es	<a href="http...	Aprendiendo Analítica de Datos □	?	?	0	
es	<a href="http...	RT @alexaceron17: #enclasesdeBigdataEpn con mi com...	?	?	1	
es	<a href="http...	#enclasesdeBigdataEpn con @alexaceron17 y la profe @...	?	?	1	
es	<a href="http...	RT @FelipeA14767468: ya manden a las casas #enclase...	?	?	1	

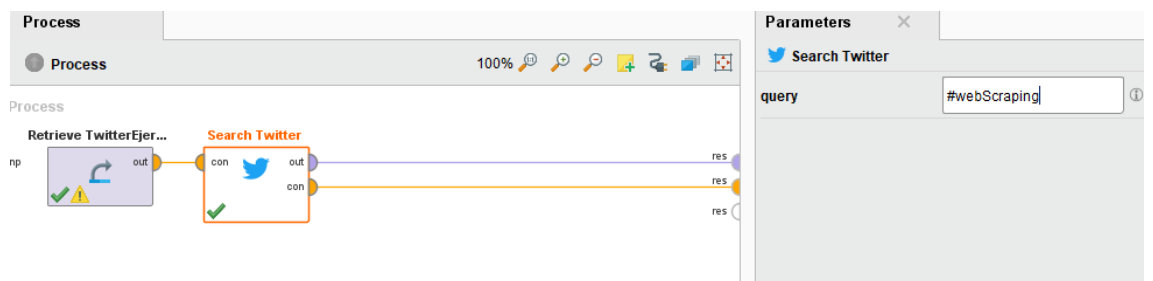


ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

- Realizar el mismo ejercicio con los Hashtags de su proyecto.



Primero busco la clave #webScraping y mi Twitt aparece en primera posición



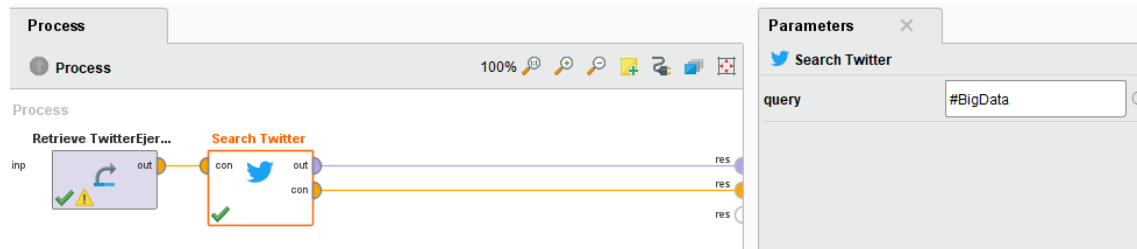
Open in Turbo Prep Auto Model Filter (100 / 100 examples)

	Created-At	From-User	From-User-Id	T...	T...	...	Sou...	Text
378...	Nov 6, 2019 6...	Danny Díaz	1192218865...	?	-1	es	<a h...	Mi proyecto trata del análisis de dat...
107...	Nov 6, 2019 6...	Gamer Geek	316331833	?	-1	en	<a h...	RT @htmlcss_stack: Parsing data f...
477...	Nov 6, 2019 5...	HTML & CSS ...	9239189187...	?	-1	en	<a h...	Parsing data from a table on a web...
278...	Nov 6, 2019 5...	#ComputeHo...	230488899	?	-1	en	<a h...	RT @jubarrosor: #WebScraping Bo...
545...	Nov 6, 2019 5...	Juan P Barro...	9888803316...	?	-1	en	<a h...	#WebScraping Board Game Descri...
167...	Nov 6, 2019 1...	ScrapeHero	2544416664	?	-1	en	<a h...	Python Frameworks and Libraries f...
437...	Nov 6, 2019 1...	Pamela Bailey	1192112274...	?	-1	en	<a h...	I will gather information for yellow p...
390...	Nov 6, 2019 1...	Pamela Bailey	1192112274...	?	-1	en	<a h...	I will do gather any scraping data.if ...
333...	Nov 6, 2019 1...	Pamela Bailey	1192112274...	?	-1	en	<a h...	I will do scrape public data from any...
128...	Nov 6, 2019 1...	Pamela Bailey	1192112274...	?	-1	en	<a h...	I Will Do Web Crawling Or Web Scr...
393...	Nov 6, 2019 1...	Pamela Bailey	1192112274...	?	-1	en	<a h...	Are you looking for Crawling Or We...
371...	Nov 6, 2019 9...	Oxylabs	9543382068...	?	-1	en	<a h...	According to Ringlead statistics, 85...
399...	Nov 6, 2019 8...	Harun Or Ro...	987327205	?	-1	en	<a h...	Check out my Gig on Fiverr: do #dat...
377...	Nov 6, 2019 8...	Adam Jorge	1020106412...	?	-1	en	<a h...	Web scraping at large scale, faster ...



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Para #BigData mi Twitt está más abajo en la lista, lo que demuestra que es un hashtag muy usado.



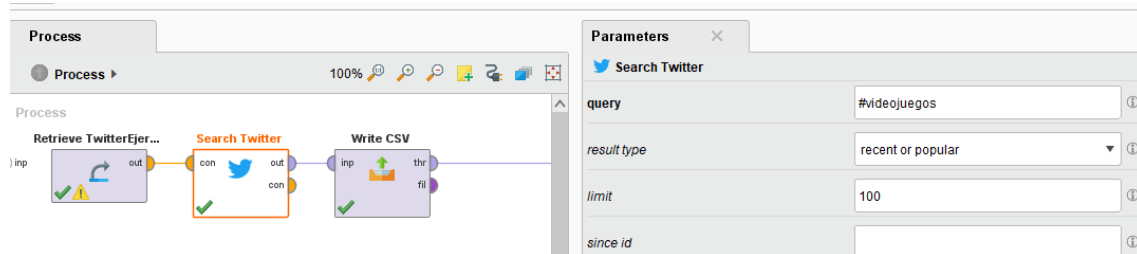
Open in [Turbo Prep](#) [Auto Model](#) Filter (100 / 100 examples): all

Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text	Gr
Jov 6, 2019 6...	Serverless Bot	1041367248...	?	-1	en	<a href="http...	RT @gp_pulipaka: 7 In...	?
Jov 6, 2019 6...	Daniel Carval...	25015865	?	-1	en	<a href="http:/...	RT @ipfconline1: Why ...	?
Jov 6, 2019 6...	ArkangelScrap	204010235	?	-1	en	<a href="http:/...	RT @SpirosMargaris: ...	?
Jov 6, 2019 6...	Britannia PR	1623145514	?	-1	en	<a href="http...	This interesting map s...	?
Jov 6, 2019 6...	Ruo Ando	320602295	?	-1	en	<a href="http:/...	id::119222487149907...	?
Jov 6, 2019 6...	Fingo	7456473692...	?	-1	en	<a href="http:/...	RT @wiomax: Unlocki...	?
Jov 6, 2019 6...	IoT and More	137405952	?	-1	en	<a href="http...	RT @SmesGoals: HT: ...	?
Jov 6, 2019 6...	CASHLess S...	18375963	?	-1	en	<a href="http...	The Landscape of #Bi...	?
Jov 6, 2019 6...	Danny Díaz	1192218865...	?	-1	es	<a href="http...	Mi proyecto trata del an...	?
Jov 6, 2019 6...	Jogo InChat(o)	127347977	?	-1	en	<a href="http...	RT @gp_pulipaka: 7 In...	?
Jov 6, 2019 6...	Jeffrey A. Willi...	1140725219...	?	-1	en	<a href="http...	RT @dinisguarda: The...	?
Jov 6, 2019 6...	Cyber Securit...	1131854274...	?	-1	en	<a href="http...	RT @ConstructechMa...	?
Jov 6, 2019 6...	Nand & Nor	741344448	?	-1	en	<a href="http...	RT @Softnet_Search: ...	?
Jov 6, 2019 6...	べてい	8451850427...	?	-1	en	<a href="http...	RT @EXAGolo: More th...	?



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Y sobre videojuegos, aunque no es una palabra que haga “string matching” directamente con mi Twitt:



ExampleSet (Search Twitter)

Open in Turbo Prep Auto Model

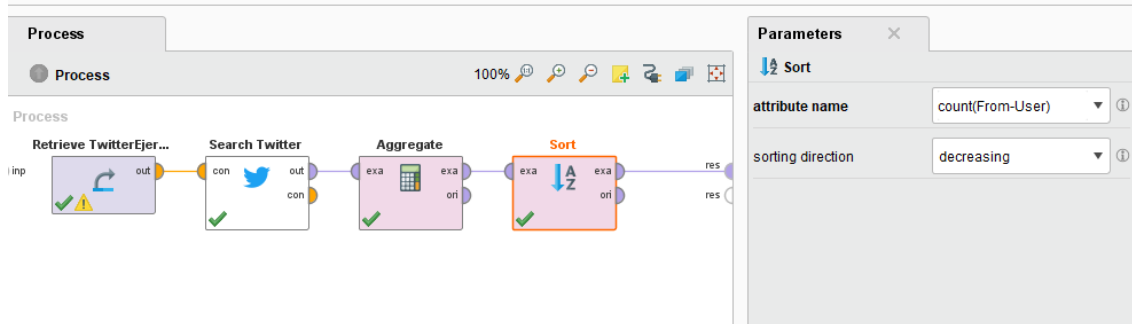
Filter (100 / 100 examples): all

Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
Nov 5, 2019 2...	Netlife Ecuador	214977276	?	-1	es	<a href="http...	Internet Archive es el lugar ideal para los verdaderos #gamers. La ...
Nov 6, 2019 4...	Ministra TIC - ...	1022259818...	?	-1	es	<a href="http:/...	El presidente @IvanDuque nos acompaña en la apertura de #Colo...
Nov 4, 2019 2...	BANDAI NAM...	2981109761	?	-1	es	<a href="http...	En un día como hoy de 2010, comenzó una saga de videojuegos q...
Nov 6, 2019 7...	AJVG	1191857590...	?	-1	es	<a href="http...	RT @odarma_mx: Jajaja yo 100% 🤪   #meme #minecraft #videoju...
Nov 6, 2019 7...	Zona de Cont...	1191866437...	?	-1	es	<a href="http...	Nuestro #blog ya está listo! #Ahora a #publicarcontenido!
Nov 6, 2019 7...	TechNoticias	1059918619...	?	-1	es	<a href="http...	La nueva solución de inteligencia artificial de Aplausos ayuda a ab...
Nov 6, 2019 7...	Javier	7725951341...	?	-1	es	<a href="http:/...	RT @NetlifeEcuador: Internet Archive es el lugar ideal para los ver...
Nov 6, 2019 7...	Sebastian Ja...	144391600	?	-1	es	<a href="http...	RT @ConstainSylvia: El presidente @IvanDuque nos acompaña e...
Nov 6, 2019 6...	Jonathan [JN...	64739568	?	-1	es	<a href="http:/...	#FacebookGaming = https://t.co/mgRh9b42Vz
Nov 6, 2019 6...	AKI Games a...	9942977675...	?	-1	en	<a href="http...	The Last Hope Free Mobile Game #VideoJuegos #VideoGames #J...
Nov 6, 2019 6...	Erick Nicolás ...	9982974332...	?	-1	es	<a href="http...	RT @BandaiNamcoLA: ¡Etiqueta a quienes siempre te acompaña...
Nov 6, 2019 6...	Microreviews	1163098468...	?	-1	es	<a href="http:/...	RT @Microreviews1: Hoy en 60 segundos... seguimos con MONK...
Nov 6, 2019 6...	Gaming RTs	7649289296...	?	-1	es	<a href="http...	RT @eSportsTheGamer: #Hearthstone   El descenso de los Drag...
Nov 6, 2019 6...	The Gamer	7440993958...	?	-1	es	<a href="http...	#Hearthstone   El descenso de los Dragones llegará a Hearthstone



**ESCUELA POLITÉCNICA NACIONAL**  
**FACULTAD DE INGENIERÍA DE SISTEMAS**  
**INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

- **Identificar el usuario más activo**



El más activo es Pedro Cuasqui. Para saber esto usamos un conteo de tweets de todos los listados

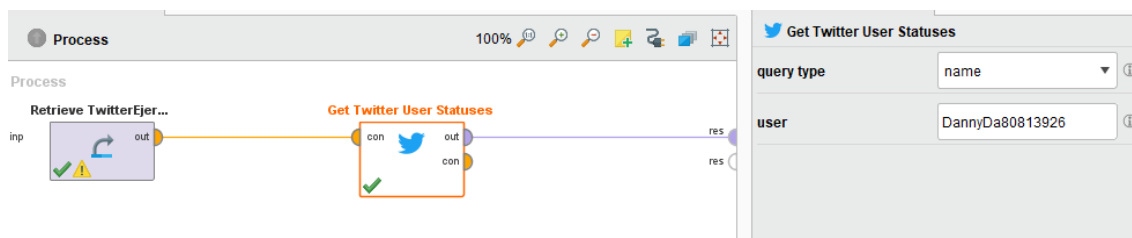
Open in Turbo Prep Auto Model

Row No.	From-User	count(From-...
1	Pedro Cuasqui	10
2	Laura Cerón	7
3	Analy Chinac...	2
4	Danny Díaz	2
5	Felipe Andrade	2

- **Utilizar los otros operadores de twitter en los cuáles se buscará por usuario, detalles de usuario y por seguidores de un usuario.**

### Por usuario

En este caso invoco a mi propio usuario





ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Y muestra los Twitts que publiqué

ExampleSet (Get Twitter User Statuses)

Open in [Turbo Prep](#) [Auto Model](#) Filter (2 / 2 examples): all

R...	Id	Created-At	From-User	Fro...	...	To...	L...	So...	Text	G...
1	11...	Nov 6, 2019 6...	Danny Díaz	1192...	?	-1	es	<a ...	Mi proyecto trata del análisis de datos de los videojuegos ...	?
2	11...	Nov 6, 2019 6...	Danny Díaz	1192...	?	-1	es	<a ...	Aquí hackeando a la NASA con html con la guía de #encla...	?

Por detalle de usuario

Invoco de nuevo mi usuario y me da el metadata correspondiente.

Process

100%

Process

Retrieve TwitterEjer...

Get Twitter User Details

Parameters

Get Twitter User Details

query type: name

user: DannyDa80813926

Open in [Turbo Prep](#) [Auto Model](#) Filter (1 / 1 examples): all

Name	ScreenName	Description	URL	Created-At	Location	Verified	Protected	Followers	Friends	Favourites	T
Danny Díaz	DannyDa808...		?	Nov 6, 2019 6...		false	false	0	0	0	2

Por seguidores de usuario

Utilizo el nombre de un usuario random de twitter

Process

100%

Process

Retrieve TwitterEjer...

Get Twitter Relations

Parameters

Get Twitter Relations

query type: name

user: william1523

relation: friends



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

---

Y obtengo una lista de sus amigos debido a que rapidMiner bloqueaba la consulta de sus seguidores con la premisa que, tenemos restringida cuota para las consultas.

Open in



Turbo Prep



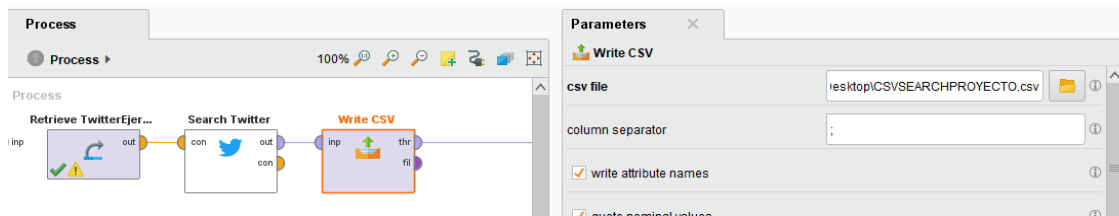
Auto Model

Row No.	Name	Friend-Id
1	william1523	16589206
2	william1523	1067858525...
3	william1523	203106197
4	william1523	2525848868
5	william1523	778564176
6	william1523	2499422425
7	william1523	218599604
8	william1523	326344160
9	william1523	385473817
10	william1523	809273
11	william1523	248409578
12	william1523	247290681
13	william1523	783214
14	william1523	14872237
15	william1523	74251501



ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

- Exportar la data encontrada a Excel o CSV



El resultado es el siguiente

Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text	Geo-Location	Geo-Location	Retweet-Count	Id
2018-11-18T19:18:18.000Z	Netlife Ecuador	214977276			-1 es	<a href="http://Internet Archive es el lugar ideal para l				20	1.1918E+18
2018-11-18T19:22:18.000Z	Ministra TIC	1.0223E+18			-1 es	<a href="http://El presidente @IvanDuque nos acompa				14	1.1922E+18
2018-11-18T19:14:18.000Z	BANDAI NAMCO	2981109761			-1 es	<a href="http://En un día como hoy de 2010, comenzó				18	1.1914E+18
2018-11-18T19:22:18.000Z	AJVG	1.1919E+18			-1 es	<a href="http://RT @odarma_mx: Jajaja yo 100% ?   #m				3	1.1922E+18
2018-11-18T19:22:18.000Z	Zona de Con	1.1919E+18			-1 es	<a href="http://Nuestro				0	1.1922E+18
2018-11-18T19:22:18.000Z	TecNoticias	1.0599E+18			-1 es	<a href="http://La nueva solución de inteligencia artifi				0	1.1922E+18
2018-11-18T19:22:18.000Z	Javier	7.726E+17			-1 es	<a href="http://RT @NetlifeEcuador: Internet Archive				20	1.1922E+18
2018-11-18T19:22:18.000Z	Sebastian Ja	144391600			-1 es	<a href="http://RT @ConstainSylvia: El presidente @iv				14	1.1922E+18
2018-11-18T19:22:18.000Z	Jonathan [ J	64739568			-1 es	<a href="http://FacebookG #####				0	1.1922E+18
2018-11-18T19:22:18.000Z	AKI Games a	9.943E+17			-1 en	<a href="http://The Last Hope Free Mobile Game #Vid				0	1.1922E+18
2018-11-18T19:22:18.000Z	Erick Nicolás	9.983E+17			-1 es	<a href="http://RT @BandaiNamcoLA: ¡Etiqueta a quie				2	1.1922E+18
2018-11-18T19:22:18.000Z	Microreview	1.1631E+18			-1 es	<a href="http://RT @Microreviews1: Hoy en 60 segund				43	1.1922E+18
2018-11-18T19:22:18.000Z	Gaming RTs	7.6493E+17			-1 es	<a href="http://RT				1	1.1922E+18
2018-11-18T19:22:18.000Z	The Gamer	7.441E+17			-1 es	<a href="http://Hearthsto				1	1.1922E+18
2018-11-18T19:22:18.000Z	Miguel Moni	44916666			-1 es	<a href="http://RT @NetlifeEcuador: Internet Archive				20	1.1922E+18
2018-11-18T19:22:18.000Z	??Bambuú-C	1392374258			-1 es	<a href="http://Conoce la nueva plataforma de crédito				0	1.1922E+18
2018-11-18T19:22:18.000Z	1912136316				-1 es	<a href="http://RT @Microreviews1: Hoy en 60 segund				43	1.1922E+18
2018-11-18T19:22:18.000Z	OFX GAMERS	4067533503			-1 es	<a href="http://Estuvimos				0	1.1922E+18
2018-11-18T19:22:18.000Z	Álvaro Varor	4779511			-1 es	<a href="http://Si que se				0	1.1922E+18
2018-11-18T19:22:18.000Z	Gallowmere	514339554			-1 es	<a href="http://RT @Microreviews1: Hoy en 60 segund				43	1.1922E+18
2018-11-18T19:22:18.000Z	Lufreaky	1.1899E+18			-1 es	<a href="http://No te				0	1.1922E+18
2018-11-18T19:22:18.000Z	TecNoticias	1.0599E+18			-1 es	<a href="http://La oferta Target Black Fridav Pixel 4 ofr				0	1.1922E+18





ESCUELA POLITÉCNICA NACIONAL  
FACULTAD DE INGENIERÍA DE SISTEMAS  
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

---

### **Análisis de resultados:**

Esta es una gran herramienta, pero tiene limitaciones de consultas, y al querer consultar cuentas con muchos seguidores el programa arrojaba un error.

Las consultas se parecen bastante al lenguaje SQL tradicional, pero estructurado por capas, haciéndolo muy fácil e incluso un inexperto en el tema podría entenderlo.

Y por último, el contenido principal de cada resultado era un texto por lo que solo podemos sacar valor si contamos con herramientas que analicen el lenguaje natural. Sin embargo, también podemos ver relaciones de influencia entre personas alrededor de un tópico puntual concentrado en un **hashtag**.

### **Conclusiones y recomendaciones:**

- RapidMiner trabaja por extracción de la información por medio de “capas” que filtran y preprocesan la consulta.
- La herramienta permite realizar procesos de Extracción, transformación y carga.
- El análisis que realiza la herramienta de Twitter es en tiempo real.
- Se realizó una introducción sobre minería de datos en los temas de “paroNacionalEc” y de los relacionados a nuestro proyecto.
- Se instaló y familiarizó con la herramienta de forma propicia para realizar los ejercicios.
- Se logró conectar con Twitter para extraer información de tweets, pero con permisos limitados.
- Se recomienda utilizar palabras claves de “moda” y palabras claves de “cola larga” (poco usadas) para expandir el análisis sobre un tópico determinado.
- Es recomendable realizar consultas sobre un conjunto pequeño de datos para entender su dominio y luego realizar procesamiento sobre N datos.

### **Referencia:**

- [1] J. L. Abellán, «Economipedia,» [En línea]. Available:  
<https://economipedia.com/definiciones/mineria-de-datos.html>. [Último acceso: 12 11 2019].
- [2] MartechForum, «MartechForum,» [En línea]. Available:  
<https://www.martechforum.com/herramienta/rapidminer/>. [Último acceso: 12 11 2019].