



Escuela Politécnica
Nacional

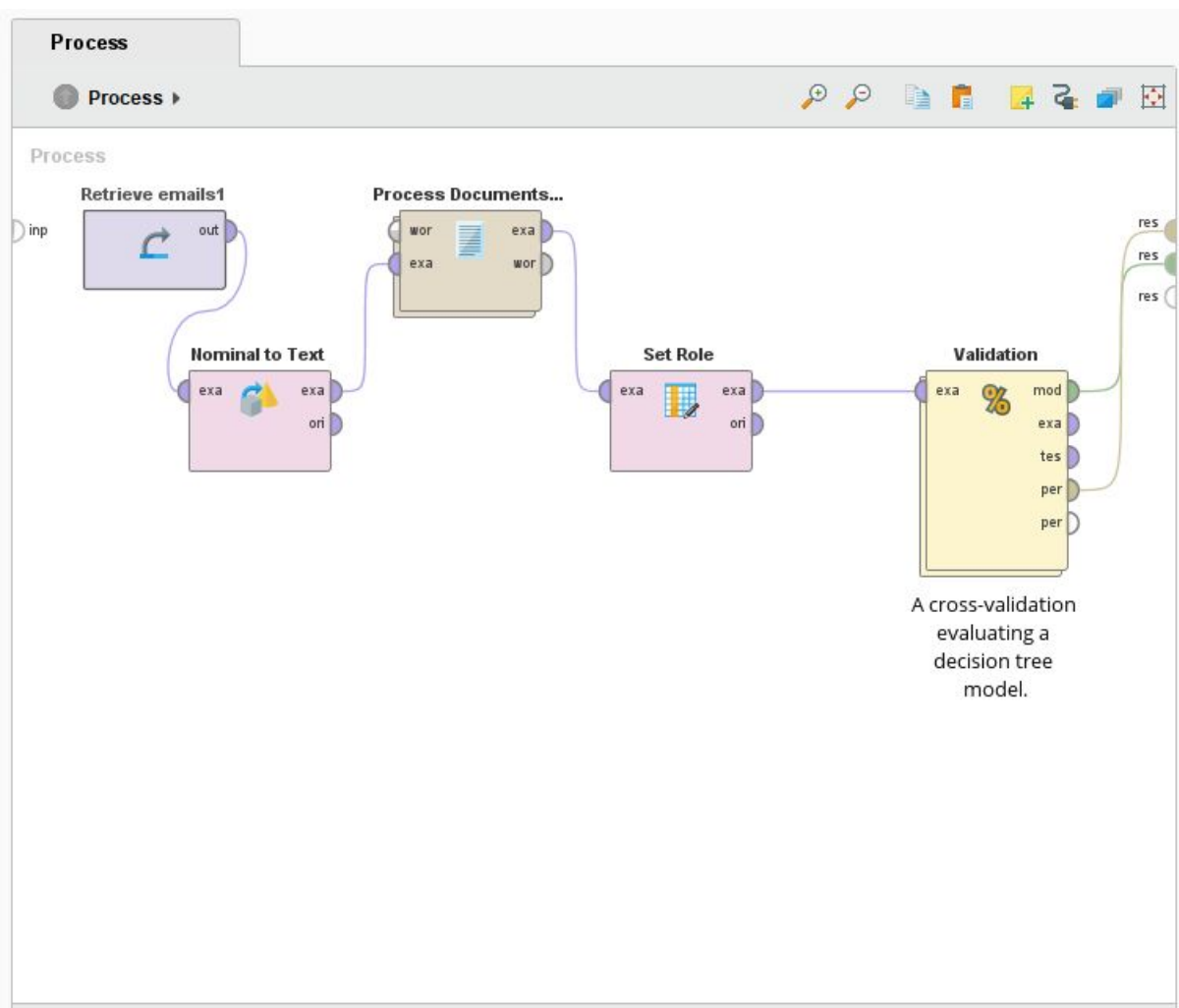
ESCUELA POLITECNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

INTELIGENCIA ARTIFICIAL - SIC 524

Gabriela Talavera
21 de agosto de 2020

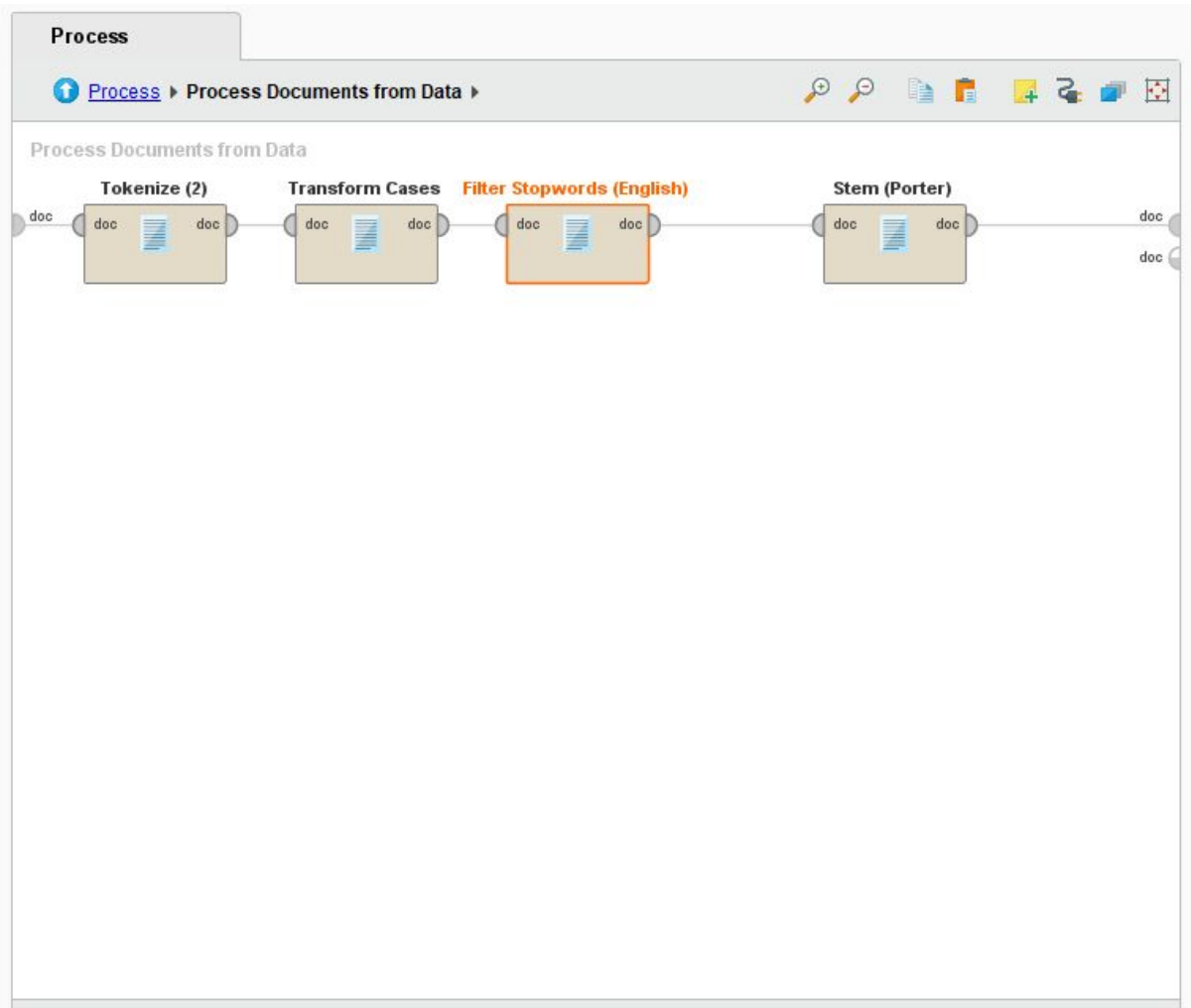
Bag of words Clasificación de SPAM y HAM con RapidMiner

Se realiza el proceso de conversión y tratamiento de datos para la clasificación de los mismos

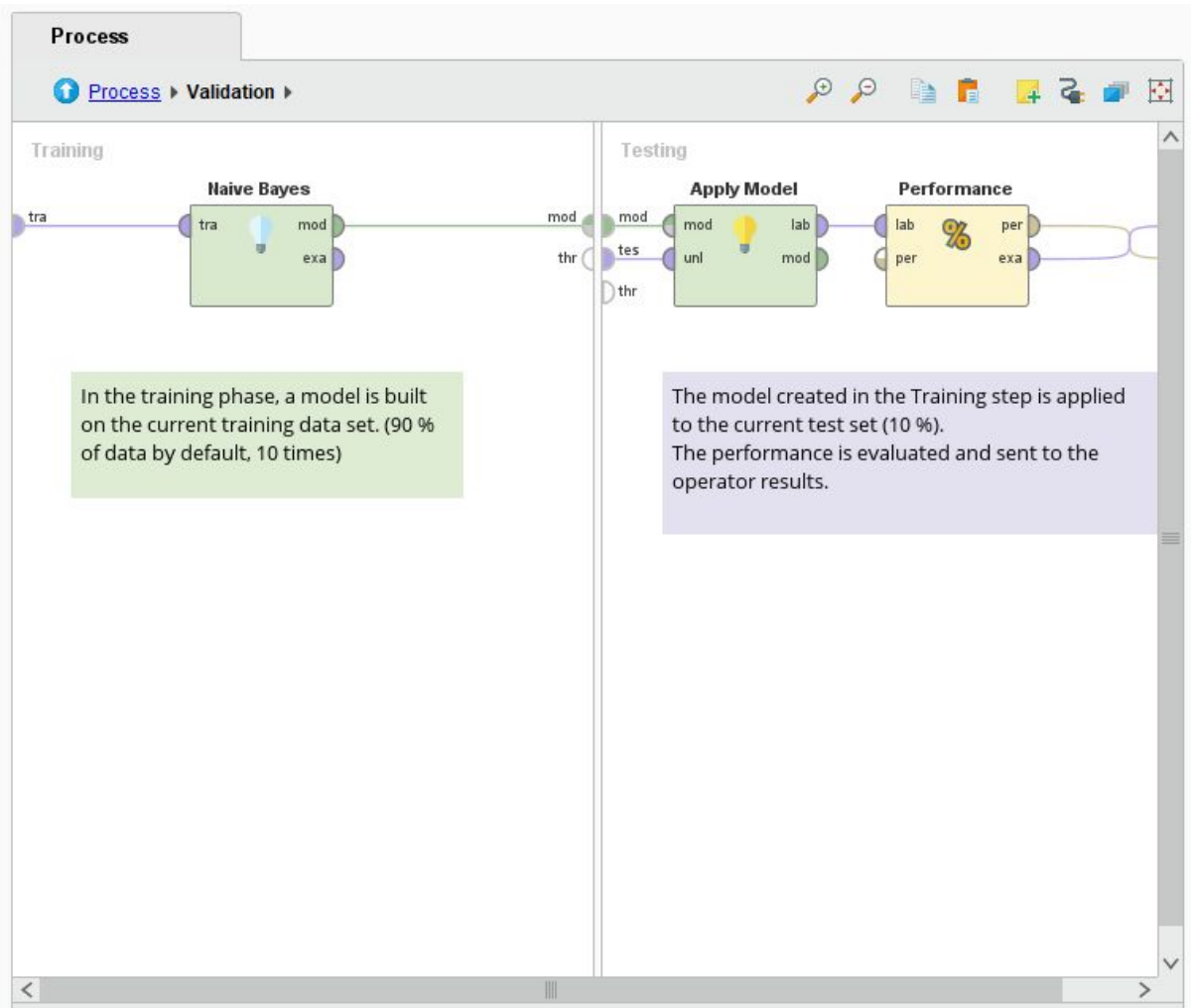




Se realiza el proceso interno para procesamiento de documentos desde los datos



Se configura la clasificación del validador de datos y la salida en forma de Bayes





Finalmente se obtiene el resultado del proceso realizado, indica que se ha ejecutado como una distribución de bayes simple.

Result History | SimpleDistribution (Naive Bayes) | PerformanceVector (Performance)

SimpleDistribution

Distribution model for label attribute Class

Class SPAM (0.230)
24335 distributions

Class HAM (0.767)
24335 distributions

Simple Charts

Distribution Table

Annotations

Además, se observan los valores de precisión y exactitud que se tiene para diferentes valores, incluyendo falsos positivos que no representan un valor significativo respecto a la cantidad de datos que se maneja.

Result History | SimpleDistribution (Naive Bayes) | PerformanceVector (Performance)

Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

Table View | Plot View

accuracy: 96.71% +/- 0.56% (micro average: 96.71%)

	true SPAM	true HAM	class precision
pred. SPAM	1133	53	95.53%
pred. HAM	127	4151	97.03%
class recall	89.92%	98.74%	



Para la precisión se observan los valores de predicción para las dos clases.

Result History

SimpleDistribution (Naive Bayes)

PerformanceVector (Performance)

Table View Plot View

precision: 97.03% +/- 0.54% (micro average: 97.03%) (positive class: HAM)

	true SPAM	true HAM	class precision
pred. SPAM	1133	53	95.53%
pred. HAM	127	4151	97.03%
class recall	89.92%	98.74%	

Criterion

accuracy

precision

recall

AUC (optimistic)

AUC

AUC (pessimistic)

Performance

Description

Annotations

Result History

SimpleDistribution (Naive Bayes)

PerformanceVector (Performance)

Table View Plot View

recall: 98.74% +/- 0.54% (micro average: 98.74%) (positive class: HAM)

	true SPAM	true HAM	class precision
pred. SPAM	1133	53	95.53%
pred. HAM	127	4151	97.03%
class recall	89.92%	98.74%	

Criterion

accuracy

precision

recall

AUC (optimistic)

AUC

AUC (pessimistic)

Performance

Description

Annotations



También se observan gráficas que evidencian el cross validation realizado a lo largo del proceso







PerformanceVector

PerformanceVector:

accuracy: 96.71% +/- 0.56% (micro average: 96.71%)

ConfusionMatrix:

True:	SPAM	HAM
SPAM:	1133	53
HAM:	127	4151

precision: 97.03% +/- 0.54% (micro average: 97.03%) (positive class: HAM)

ConfusionMatrix:

True:	SPAM	HAM
SPAM:	1133	53
HAM:	127	4151

recall: 98.74% +/- 0.54% (micro average: 98.74%) (positive class: HAM)

ConfusionMatrix:

True:	SPAM	HAM
SPAM:	1133	53
HAM:	127	4151

AUC (optimistic): 0.993 +/- 0.002 (micro average: 0.993) (positive class: HAM)

AUC: 0.733 +/- 0.214 (micro average: 0.733) (positive class: HAM)

AUC (pessimistic): 0.890 +/- 0.020 (micro average: 0.890) (positive class: HAM)