

**Escuela de Ingeniería en Computación**

Electiva  
Recuperación de la Información Textual

**Proyecto programado III  
Clasificadores**

**Profesor:**

Jose Enrique Araya Monge

**Estudiantes:**

Natalia Vargas Reyes  
Javier Rivera Madrigal

**Fecha de entrega:**

18/1/2021

**II Semestre 2021**

## Introducción

La tarea consiste en aplicar y evaluar los clasificadores Rocchio y bayesianos ingenuos, donde se utilizan dos vectores, uno de entrenamiento y uno de prueba, que cada línea del vector es un documento con información importante el identificador, la clase a la que pertenece el documento, el total de términos en el mismo y estos términos que aparecen con su respectivo peso.

En el caso del clasificador Rocchio se van a usar tres pares de parámetros diferentes, variando el  $\beta$  y  $\gamma$ . Al finalizar cada clasificador dando como resultado el documento que pertenece al vector de prueba con sus clases(real y asignada por el clasificador) se evaluarán los resultados usando métricas tales como: precisión, recall, acierto y error.

## Estado del proyecto

ETAPA	% DE COMPLET.	COMENTARIO O ACLARACIÓN
<b>Lectura y conversión</b>		
Puede leer y convertir datos de test_set	100%	
Puede leer y convertir datos de train_set	100%	
<b>CLASIFICADOR ROCCHIO</b>		
Calcula centroides	100%	
Calcula similitudes centroides-vectores para los tres casos	100%	
Produce un escalafón para cada documento del test_set para cada uno de los tres casos	100%	
Asignar clases a los documentos del test_set para cada uno de los tres casos	100%	
<b>CLASIFICADOR BAYESIANO INGENUO</b>		
Calcula los valores $p_{ip}$ y $q_{ip}$ para los términos	100%	

Calcula similitudes para los documentos del test_set	100%	
Produce un escalafón para cada documento del test_set	100%	
Asigna clases a los documentos del test_set	100%	
Evaluación (puede ser manual)	100%	
Elabora la tabla de contingencia para cada clase	100%	
Calcula las métricas: acierto, error, precisión y exhaustividad	100%	
<b>EVALUACIÓN (puede ser manual)</b>		
Elabora la tabla de contingencia para cada clase	100%	
Calcula las métricas: acierto, error, precisión y exhaustividad	100%	

## Evaluación clasificador rocchio

Para los dos primeros parámetros se obtuvieron resultados iguales

----parámetros-----

a = 0,75 y b = 0,25

a = 0,85 y b = 0,15

	Coffee	No Coffee	np		Acierto	0,967
Coffee	4	1	5		Error	0.0333
No Coffee	0	19	26		Recall	1
ni	4	27	31		Precisión	0.8

	money-fx	No money-fx	np		Acierto	0,96774194
money-fx	5	1	6		Error	0,033
No money-fx	0	25	25		Recall	1
ni	5	26	31		Precisión	0,833

	grain	No grain	np		Acierto	1
grain	11	0	11		Error	0
No grain	0	20	20		Recall	1
ni	11	20	31		Precisión	1

	crude	No crude	np		Acierto	1
crude	5	0	5		Error	0
No crude	0	26	26		Recall	1
ni	5	26	31		Precisión	1

	trade	No trade	np		Acierto	0,93548387
trade	4	0	4		Error	0,06451613
No trade	2	25	27		Recall	0,66666667
ni	6	25	31		Precisión	1

----parámetros-----

a = 0,95 y b = 0,05

	Coffee	No Coffee	np		Acierto	0,935
Coffee	4	2	6		Error	0.064
No Coffee	0	25	25		Recall	1
ni	4	27	31		Precisión	0.66

	money-fx	No money-fx	np		Acierto	0,96774194
money-fx	5	1	6		Error	0,033
No money-fx	0	25	25		Recall	1
ni	5	26	31		Precisión	0,833

	grain	No grain	np		Acierto	1
grain	11	0	11		Error	0
No grain	0	20	20		Recall	1
ni	11	20	31		Precisión	1

	crude	No crude	np		Acierto	1
crude	5	0	5		Error	0
No crude	0	26	26		Recall	1
ni	5	26	31		Precisión	1

	trade	No trade	np		Acierto	0,9032
trade	3	0	3		Error	0,0967
No trade	3	25	28		Recall	0,5
ni	6	25	31		Precisión	1

## Evaluación clasificador bayesianos ingenuos

Para ver la evaluación manual del clasificador bayesianos, se puede ingresar en el siguiente link para acceder a la hoja de cálculo donde se muestran las tablas respectivas de cada clase con aciertos, error, recall y precisión.

[https://docs.google.com/spreadsheets/d/1OUQZi5\\_3Ksd-w3hITqjPEYNdG7e9g\\_ZngJiz\\_D3zh1s/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1OUQZi5_3Ksd-w3hITqjPEYNdG7e9g_ZngJiz_D3zh1s/edit?usp=sharing)

También se muestran a continuación las evaluaciones.

DocId	Real	Estimado
18	coffee	coffee
52	money-fx	coffee
68	money-fx	money-fx
111	trade	trade
118	coffee	coffee
121	trade	coffee
142	grain	grain
152	grain	grain
170	crude	crude
192	crude	crude
208	grain	grain
219	grain	coffee
226	grain	coffee
235	money-fx	money-fx
260	trade	coffee
273	money-fx	money-fx
322	grain	grain
343	grain	grain
354	trade	trade
365	trade	trade
424	grain	coffee
444	coffee	coffee
445	grain	grain

463	crude	crude
477	grain	grain
479	coffee	coffee
482	grain	coffee
496	money-fx	money-fx
498	crude	coffee
506	trade	trade
519	crude	crude

	Coffe	No Coffe	np		Acierto	0,7419354 8
Coffe	4	8	12		Error	0,2580645 2
No Coffe	0	19	19		Recall	1
ni	4	27	31		Precisión	0,3333333 3

	money-fx	No money-fx	np		Acierto	0,9677419 4
money-fx	4	0	4		Error	0,0322580 6
No money-fx	1	26	27		Recall	0,8
ni	5	26	31		Precisión	1



	trade	No trade	np		Acierto	0,93548387
trade	4	0	4		Error	0,06451613
No trade	2	25	27		Recall	0,66666667
ni	6	25	31		Precisión	1

	grain	No grain	np		Acierto	0,87096774
grain	7	0	7		Error	0,12903226
No grain	4	20	24		Recall	0,63636364
ni	11	20	31		Precisión	1

	crude	No crude	np		Acierto	0,96774194
crude	4	0	4		Error	0,03225806
No crude	1	26	27		Recall	0,8
ni	5	26	31		Precisión	1

## **Comentarios finales**

Para el estado final del programa, la parte de lectura y conversión quedó excelente, ya que obtiene los datos de forma correcta. Los clasificadores cumplen su función y devuelve el escalafón con las clases asignadas por el mismo, esta parte de hecho la hace mediante un documento de texto y la asignación de clases sólo se imprime.

Además para cada clasificador se llegó a la conclusión de que los valores estimados son frecuentemente acertados, lo que podría indicar que los cálculos están bien hechos.

En general, se lograron realizar todos los puntos solicitados, con un funcionamiento adecuado.