

# **PRÁCTICA 3:**

## **CLASIFICADORES SIMPLES DE DOCUMENTOS REALES**

**MINERÍA DE DATOS**  
**MIGUEL CHAVEINTE GARCÍA**

## 1. EJERCICIO 1

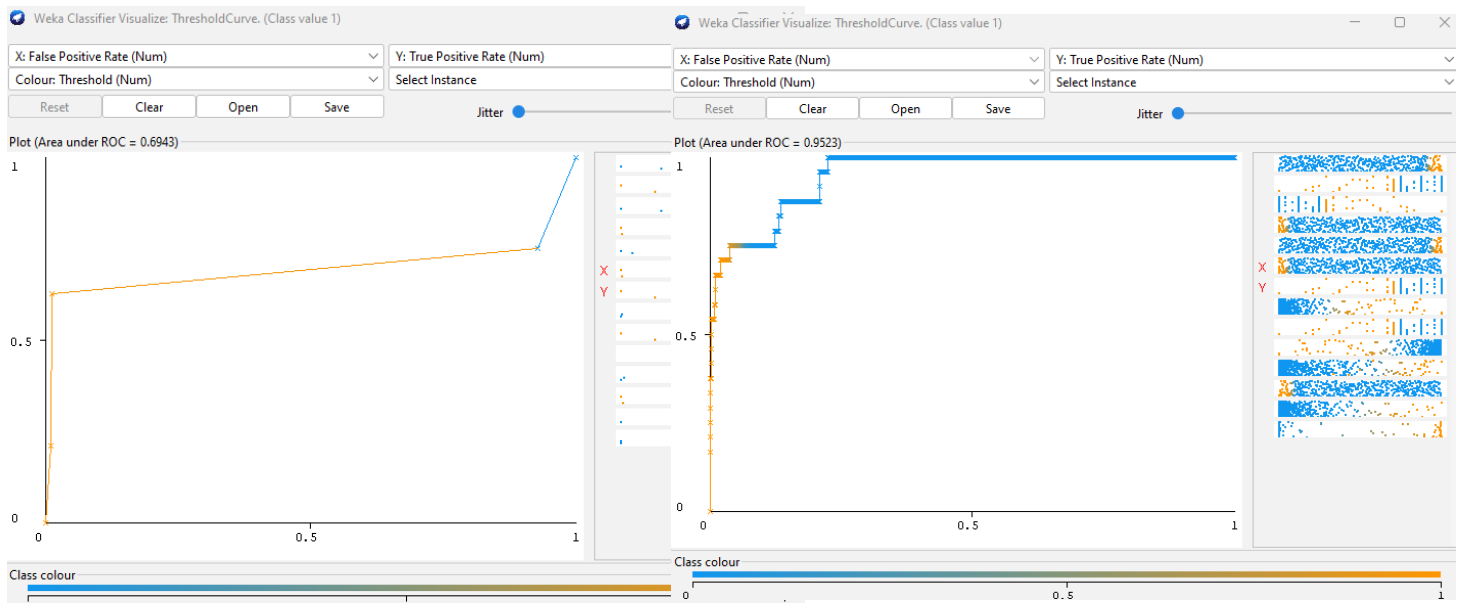
PORCENTAJES DE ACIERTOS SOBRE LOS CONJUNTOS DE TEST		
	J48	NBMultinomial
Corn	97.351%	93.709%
Grain	96.358 %	90.729%

El J48 se consiguen mejores tasas de acierto que el NBMultinomial para estos conjuntos de datos.

TASAS DE ERROR SOBRE LOS CONJUNTOS DE TEST		
	J48	NBMultinomial
Corn	0.027	0.063
Grain	0.036	0.093

Si observamos las tasas de error, elegiría el algoritmo J48 ya que llegamos a obtener tasas de error muchos menores, incluso, aproximadamente la mitad que para NBMultinomial.

## 2. EJERCICIO 2



Estas gráficas de las curvas ROC para CORN corresponden a J48 Y NBMultinomial, de izquierda a derecha respectivamente. Podemos ver como el área bajo la curva ROC de J48 es 0.694 y la de NBMultinomial es 0.952, por lo que este último funciona mejor para Corn si atendemos al área bajo la curva.

Para el caso de Grain tenemos un 0.906 para J48 y 0.973 para NBMultinomial. Por lo que al igual que en el caso anterior NB obtiene un resultado mejor, aunque es más ajustado que en el caso anterior.

### 3. EJERCICIO 3

- Precision: Porcentaje de las instancias recuperadas que tienen relevancia ( $TP/(TP+FP)$ ). Si el valor es próximo a 1, nos indica que todos los documentos recuperados son relevantes. Cuanto menor sea el número de falsos positivos detectados, más aumentará la precisión.
- Recall: Porcentaje de las instancias relevantes que han sido recuperadas ( $tp=TP/P$ ). Si el valor se acerca a 1 indica que se recuperan los documentos relevantes. El mejor valor es aquel que el valor de los ciertos positivos sea próximo al total de positivos clasificados.
- F-measure: Medida de precisión que pondera la precisión y el recall, dando la misma importancia a ambos valores  $\rightarrow ((2 * recall * precision) / (recall + precision))$ .

CORN			
	Precision	Recall	F-measure
J48	0.682	0.625	0.652
NBMultinomial	0.36	0.75	0.486

GRAIN			
	Precision	Recall	F-measure
J48	0.966	0.995	0.98
NBMultinomial	0.99	0.907	0.947

Teniendo en cuenta estos resultados, para Corn funciona mejor el J48, aunque en recall se quede atrás.

Para Grain, también me decantaría por J48, al igual que el caso anterior el NBMultinomial gana en Precision.