**Naive Byes Classifier**

**Definición:**

La clasificación bayesiana ingenua es un tipo de clasificación bayesiana simple y probabilística basada en el teorema de Bayes con una fuerte independencia (llamada ingenua) de las hipótesis. Implementa un ingenuo clasificador bayesiano, o clasificador ingenuo bayesiano, perteneciente a la familia de los clasificadores lineales. En términos sencillos, un clasificador bayesiano ingenuo asume que la existencia de una característica para una clase es independiente de la existencia de otras características. Una fruta puede ser considerada como una manzana si es roja, redondeada y de unos diez centímetros de largo. Incluso si estas características están vinculadas en la realidad, un ingenuo clasificador bayesiano determinará que la fruta es una manzana considerando independientemente estas características de color, forma y tamaño.

**Resultados:**

El modelo bayesiano calcula la distribución gaussiana de los atributos numéricos. En nuestro caso los valores son la frecuencia de los términos: TF.

**Los parámetros:**

Porbabilidad= 0.01

|  |  |  |  |
| --- | --- | --- | --- |
| Category | precision | recall | f.mesure |
| Negative | 0.66 | 0.78 | 0.72 |
| Positive | 0.73 | 0.60 | 0.66 |
| Accuracy | 0.69 | | |

Probabilidad = 0.005

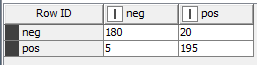
|  |  |  |  |
| --- | --- | --- | --- |
| Category | precision | recall | f.mesure |
| Negative | 0.71 | 0.80 | 0.75 |
| Positive | 0.68 | 0.77 | 0.72 |
| Accuracy | 0.74 | | |

Probabilidad = 0.001

|  |  |  |  |
| --- | --- | --- | --- |
| Category | precision | recall | f.mesure |
| Negative | 0.78 | 0.80 | 0.79 |
| Positive | 0.79 | 0.77 | 0.78 |
| Accuracy | 0.78 | | |

Probabilidad = 0

|  |  |  |  |
| --- | --- | --- | --- |
| Category | precision | recall | f.mesure |
| Negative | 0.97 | 0.9 | 0.93 |
| Positive | 0.9 | 0.97 | 0.94 |
| Accuracy | 0.93 | | |

La matriz de confusión correspondiente a los mejores resultados :

**Conclusiones:**

El clasificador de Naïve Byes es el que nos ha dado los mejores resultados para la clasificación de textos, al disminuir el valor de probabilidad aumenta la precisión y alcanza el valor de 90% de buena clasificación para documentos positivos y 97% para documentos negativos.