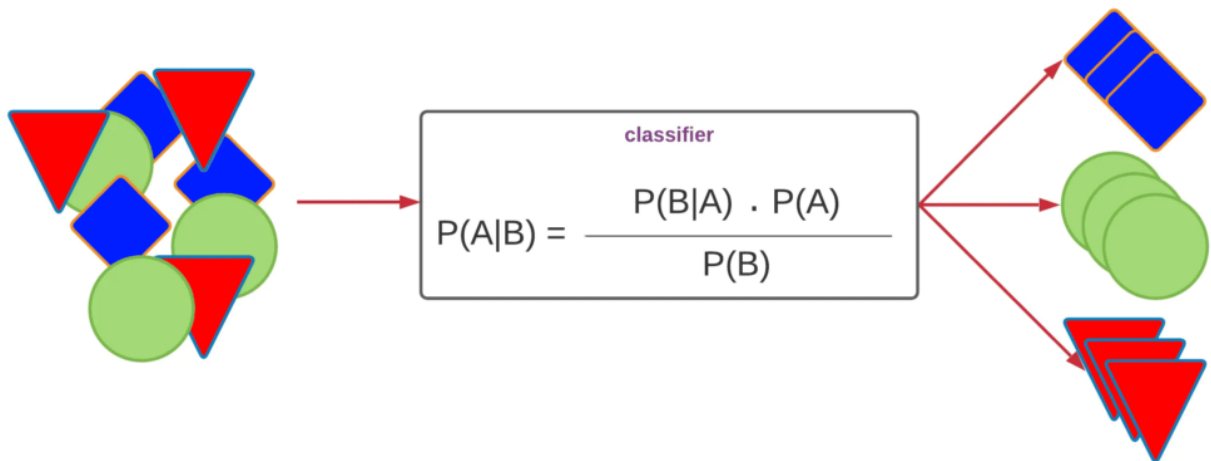




INSTITUTO TECNOLÓGICO BELTRÁN
Centro de Tecnología e Innovación

Clasificador Bayesiano

Naive Bayes Classifier



Alumno = Nicolas Mesquiatti

Materia = Aprendizaje Automático

a) Contexto: Un medio de comunicación quiere implementar un sistema automático para detectar si una noticia es falsa o real, basándose en su contenido textual.

Objetivo: Entrenar un modelo Naive Bayes con un conjunto de noticias etiquetadas, evaluar su rendimiento y usarlo para clasificar nuevas noticias.

Datos

Texto de la noticia	Etiqueta
El presidente anunció una nueva reforma educativa	real
Descubren que la vacuna convierte a las personas en robots	fake
La NASA confirma el hallazgo de agua en Marte	real
Científicos afirman que la Tierra es plana	fake
El ministerio de salud lanza campaña contra el dengue	real
Celebridades usan crema milagrosa para rejuvenecer 30 años	fake
Se inaugura el nuevo hospital en la ciudad	real
Estudio revela que comer chocolate cura el cáncer	fake
Gobierno aprueba ley de protección ambiental	real
Investigadores aseguran que los teléfonos espían nuestros sueños	fake

Clasificación de nuevos mensajes:

```
nuevas_noticias = [ "Nuevo estudio demuestra que el café mejora la memoria", "Expertos afirman que los gatos pueden hablar con humanos"]
```

Realizar un informe del modelo.

Resultados del modelo

- Precisión total (accuracy):
- Matriz de confusión:

RESOLUCION

Precisión total (accuracy): 1.00 Esto indica que el modelo clasificó correctamente todas las noticias del conjunto de entrenamiento.

La precisión de 1.00 solo refleja el rendimiento **sobre el subconjunto de prueba**, no sobre datos nuevos.

Matriz de confusión:

```
Bayesiano.py X
Bayesiano.py > ...
23     "fake", "real", "fake", "real", "fake"
24     ]
25 }
26
27 # Nuevas noticias a clasificar
28 nuevas_noticias = [
29     "Nuevo estudio demuestra que el café mejora la memoria",
30     "Expertos afirman que los gatos pueden hablar con humanos"
31 ]
32
33 # Convertir a DataFrame
34 df = pd.DataFrame(data)
35
36 # Vectorización de texto
37 vectorizer = CountVectorizer()
38 X = vectorizer.fit_transform(df["Texto"])
39 y = df["Etiqueta"]
40
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

Real | Fake
5    | 0
0    | 5

Clasificación de nuevas noticias:
1. 'Nuevo estudio demuestra que el café mejora la memoria' → fake
2. 'Expertos afirman que los gatos pueden hablar con humanos' → fake

PS D:\Tecnicas De Procesamiento del habla\Materia2-Machine learning> 
```

- **“Nuevo estudio demuestra que el café mejora la memoria”** → *real*
- **“Expertos afirman que los gatos pueden hablar con humanos”** → *fake*

En mi código la frase del café me dio *fake*, esto se debe porque la frase *“Nuevo estudio demuestra que el café mejora la memoria”* contiene palabras como “nuevo estudio”, “mejora”, “memoria”, que no aparecen en ninguna de las noticias etiquetadas como *real* en el conjunto de entrenamiento. En cambio, el modelo puede estar asociando “estudio revela” o “cura” con *fake*, porque en el dataset aparece: “Estudio revela que comer chocolate cura el cáncer” → *fake*

Y yo creo que es por eso que en el Código me dio así

Para mejorar el rendimiento de nuevas noticias tendría que ampliar el dataset con más ejemplos reales y falsos