

Klasifikasi Dokumen dengan Naive Bayes

1. Melakukan tokenisasi dan stemming pada D7

D7 = "Demo buruh menaikkan upah minimum"

Tokenisasi	"demo"	"buruh"	"menaikkan"	"upah"	"minimum"
Stemming	"demo"	"buruh"	"naik"	"upah"	"minimum"

2. Menghitung probabilitas prior

- Jumlah seluruh dokumen (N) = 6
- Jumlah dokumen dalam kategori Olahraga (N_Olahraga) = 2
- Jumlah dokumen dalam kategori Politik (N_Politik) = 2
- Jumlah dokumen dalam kategori Ekonomi (N_Ekonomi) = 2

Masukkan ke rumus $P(kategori) = \frac{N(kategori)}{N}$

- $P(Olahraga) = \frac{N(Olahraga)}{N} = \frac{2}{6} = \frac{1}{3}$
- $P(Politik) = \frac{N(Politik)}{N} = \frac{2}{6} = \frac{1}{3}$
- $P(Ekonomi) = \frac{N(Ekonomi)}{N} = \frac{2}{6} = \frac{1}{3}$

3. Menghitung probabilitas bayes

- Olahraga: $P(Olahraga | D7) = \frac{1}{3} \cdot \frac{2}{14} \cdot \frac{2}{14} \cdot \frac{1}{14} \cdot \frac{1}{14} \cdot \frac{1}{14}$
- Politik: $P(Politik | D7) = \frac{1}{3} \cdot \frac{1}{14} \cdot \frac{1}{14} \cdot \frac{1}{14} \cdot \frac{2}{14} \cdot \frac{1}{14}$
- Ekonomi: $P(Ekonomi | D7) = \frac{1}{3} \cdot \frac{1}{14} \cdot \frac{1}{14} \cdot \frac{2}{14} \cdot \frac{1}{14} \cdot \frac{2}{14}$

4. Normalisasi kategori menggunakan rumus $P(kategori | D7) = \frac{P(kategori | D7)}{P(total)}$

Normalisasi probabilitas:

- Olahraga: $P(Olahraga | D7) = \frac{2.14577 \times 10^{-7}}{(2.14577 \times 10^{-7}) + (7.79978 \times 10^{-7}) + (1.3221 \times 10^{-7})} = 0.37$
- Politik: $P(Politik | D7) = \frac{7.79978 \times 10^{-7}}{(2.14577 \times 10^{-7}) + (7.79978 \times 10^{-7}) + (1.3221 \times 10^{-7})} = 0.23$
- Ekonomi: $P(Ekonomi | D7) = \frac{1.3221 \times 10^{-7}}{(2.14577 \times 10^{-7}) + (7.79978 \times 10^{-7}) + (1.3221 \times 10^{-7})} = 0.40$

5. Kesimpulan

Karena nilai probabilitas terbesar adalah kategori Ekonomi maka dapat disimpulkan bahwa D7 termasuk ke dalam kategori Ekonomi.