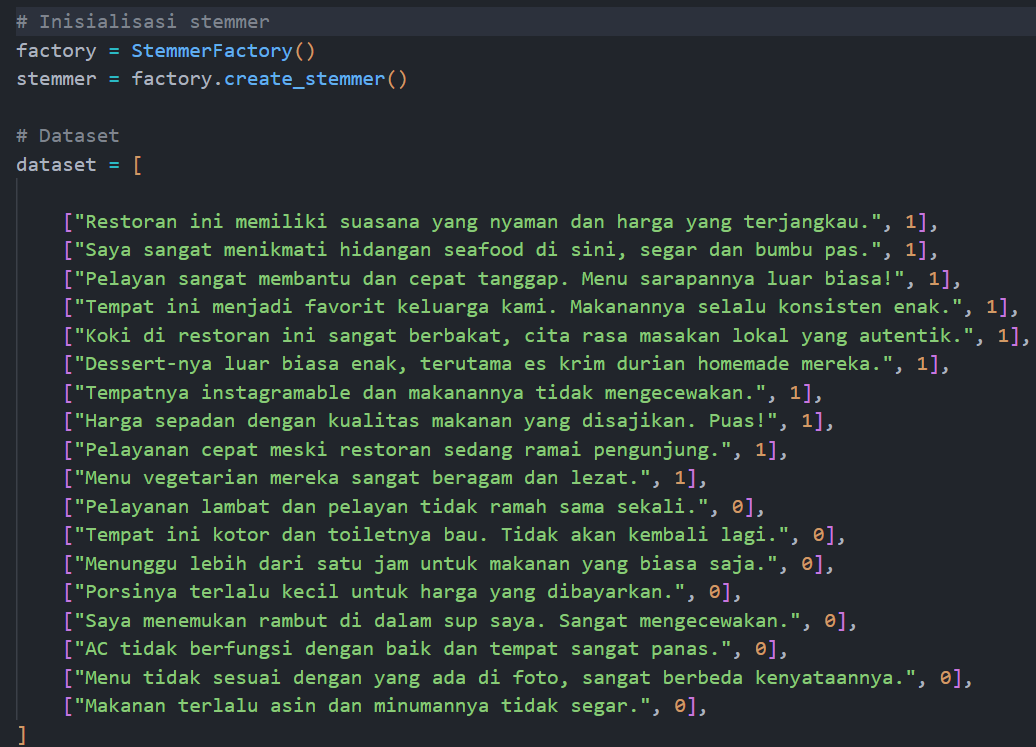
**Penggalian Sosial Media**

**Tugas 4**

**Nama : Cahya Galur Permana**

**Nim : 10221057**

1. Membuat Dataset Sendiri
   * Buat dataset teks Anda sendiri (misalnya ulasan film, produk, atau restoran) dengan label sentimen (1 untuk positif, 0 untuk negatif).
   * Latih model menggunakan dataset tersebut dan uji performanya

A screen shot of a computer program

AI-generated content may be incorrect.

A computer screen shot of text

AI-generated content may be incorrect.A screen shot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

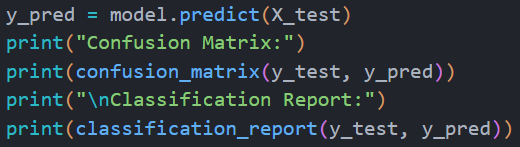
Dari 5 text baru, model dengan benar memprediksi 2 sentimen. Hasil uji performa menampilkan akurasi sebesar 75%

1. Penerapan pada Kasus Dunia Nyata
   * Cari dataset teks publik (misalnya dari Kaggle atau UCI Machine Learning Repository) yang berisi ulasan atau komentar.
   * Terapkan pipeline yang sama (CountVectorizer + MultinomialNB) untuk melakukan analisis sentimen pada dataset tersebut

Dataset yang digunakan adalah dataset ulasan aplikasi Gojek.

A screen shot of a computer program

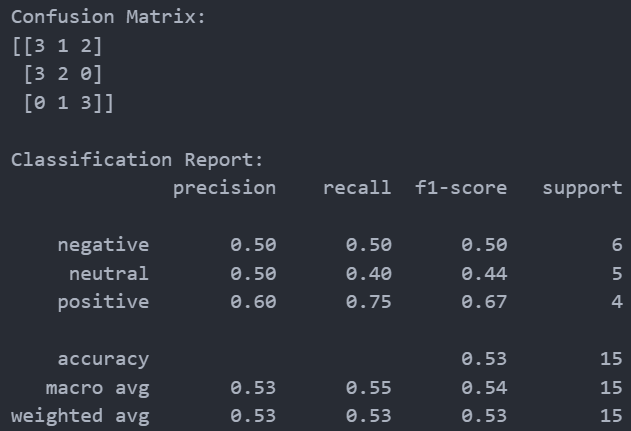
AI-generated content may be incorrect.A computer screen shot of text

AI-generated content may be incorrect.A screenshot of a computer screen

AI-generated content may be incorrect.

Dari hasil evaluasi model sentiment analysis dari dataset ulasan gojek. Model mencapai akurasi sebesar 53%. Model cukup baik dalam mengidentifikasi kelas positif, dengan nilai *recall* sebesar 0.75, artinya 75% dari data positif dikenali dengan benar. Kelas neutral memiliki performa terburuk, terutama dari segi *recall* (0.40), menandakan banyak data neutral diklasifikasi sebagai kelas lain (terutama sebagai negative). Model lebih cenderung mengklasifikasikan data ke negative dan positive, dan kurang bisa membedakan neutral.

1. Optimasi Parameter
   * Eksperimen dengan parameter alpha pada MultinomialNB (default adalah 1.0): classifier = MultinomialNB(alpha=0.5)
   * Amati bagaimana nilai alpha memengaruhi performa model Setelah merubah parameter alpha menjadi 0.5:



Dengan alpha = 1.0, model mencapai akurasi sebesar 0.53, dan hasil yang sama juga diperoleh saat alpha diturunkan menjadi 0.5, menunjukkan bahwa perubahan nilai smoothing ini tidak memberikan dampak signifikan terhadap performa keseluruhan. F1-score untuk kelas "negatif" tetap di angka 0.50, demikian juga untuk kelas "netral" yang tetap rendah di kisaran 0.44–0.45, menandakan bahwa model belum mampu mengenali pola pada kelas ini dengan baik. Kelas "positif" menunjukkan performa terbaik dengan f1-score sebesar 0.67 pada kedua pengaturan alpha, yang didorong oleh recall tinggi (0.75) dan precision sebesar 0.60.

1. Buatkan visualisasi data distribusi frekuensi kata menggunakan library seperti Matplotlib atau Seaborn dan Tambahkan Confusion Matrix

A graph of a bar graph

AI-generated content may be incorrect.

A diagram of a negative and negative matrix

AI-generated content may be incorrect.

1. Tuliskan refleksi tentang apa yang Anda pelajari dari eksperimen ini. Apa tantangan yang Anda hadapi? Bagaimana Anda mengatasinya?