Nama : Muhamad Faqih Azhar

NIM : 0110221092

Kelas : 5TI03 / Senin Malam

**Pembuatan Project Natural Language Processing (NLP)**

**Topic**

Language Identifier

**Sumber referensi:**

<https://translatedlabs.com/language-identifier>

**GitHub Link**

<https://github.com/FaqihAzh/text-summarize-nlp>

**Deskripsi Umum**

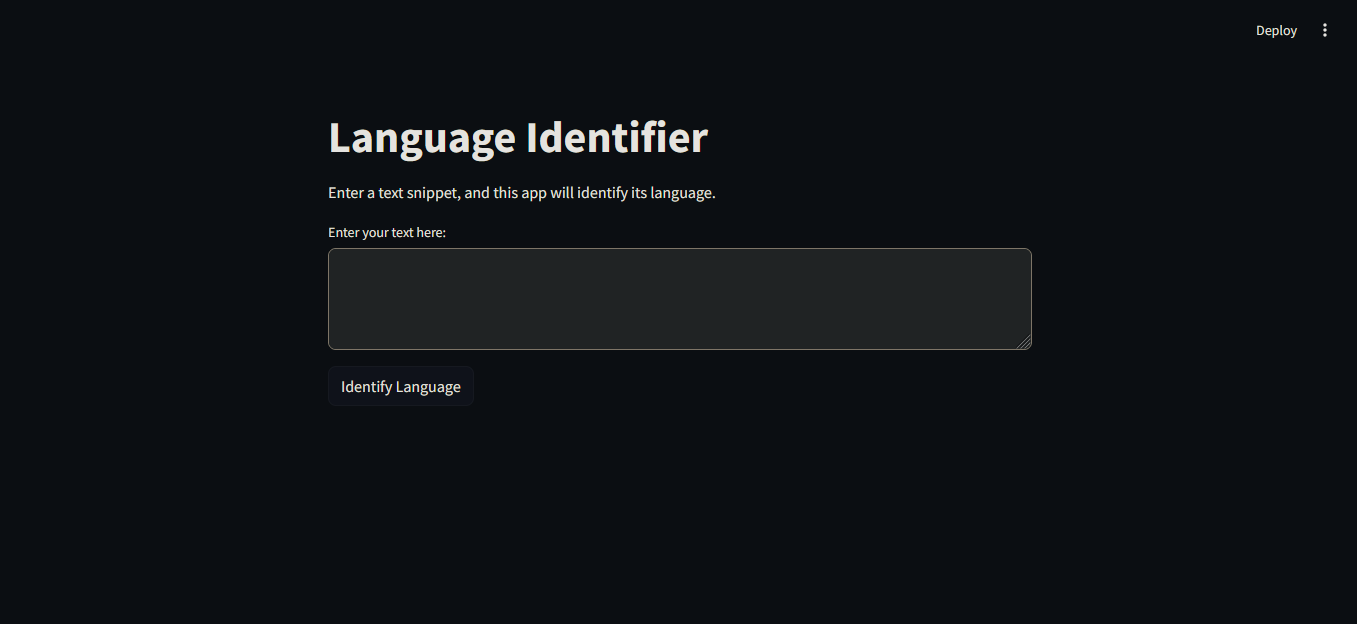
Proyek ini bertujuan untuk membuat aplikasi deteksi bahasa berbasis teks yang dapat mengidentifikasi bahasa dari teks masukan pengguna. Dengan menggunakan *library* seperti langdetect untuk pendeteksian bahasa dan streamlit untuk membuat antarmuka pengguna berbasis *web.* Apa itu *library* langdetect? library langdetect, merupakan implementasi *open-source* dari deteksi bahasa berbasis algoritma *language detection* Google, seperti yang digunakan dalam Google Translate. *Library* ini dirancang untuk mengenali lebih dari 50 bahasa berdasarkan analisis pola teks masukan. Langdetect bekerja dengan cara:

1. Membagi teks menjadi unit-unit kecil (*ngrams*), yang merupakan potongan kata atau kombinasi huruf.
2. Membandingkan pola *ngrams* ini dengan model probabilistik untuk setiap bahasa yang telah dilatih sebelumnya.
3. Menghasilkan kode bahasa dalam format standar ISO 639-1 (misalnya, id untuk Bahasa Indonesia, en untuk Bahasa Inggris).

Selain mendeteksi bahasa utama, *library* ini juga mampu memberikan daftar probabilitas untuk bahasa yang mungkin sesuai dengan teks masukan.

**Laporan Akhir**

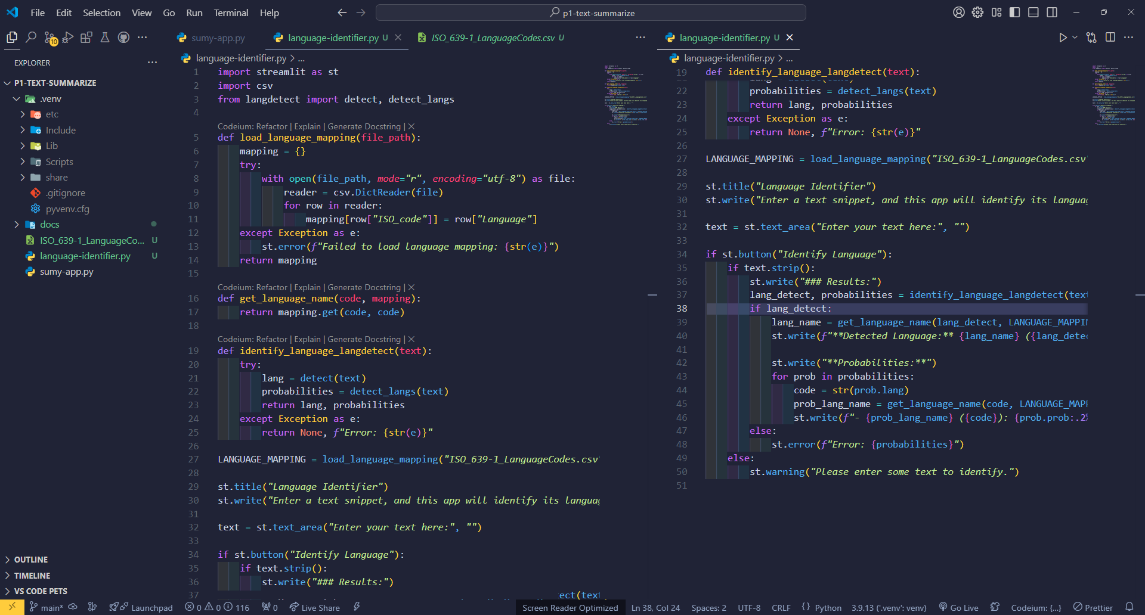
Berikut ini adalah hasil akhir dari project *Language Identifier App* yang saya buat:



Gambar 1 Tampilan Language Identifier App

Pada gambar 1 di atas, merupan tampilan akhir dari *project* yang saya kerjakan. Pada halaman tersebut, nantinya pengguna dapat memasukan teks yang akan di deteksi, setelahnya ketika pengguna menekan tombol identifikasi, maka sistem otomatis akan memproses dan mengidentifikasi bahasa yang di *input* tersebut. Bagaimana tahapan proses pengerjaannya sampai menjadi hasil akhir seperti ini? Prosesnya sebagai berikut:

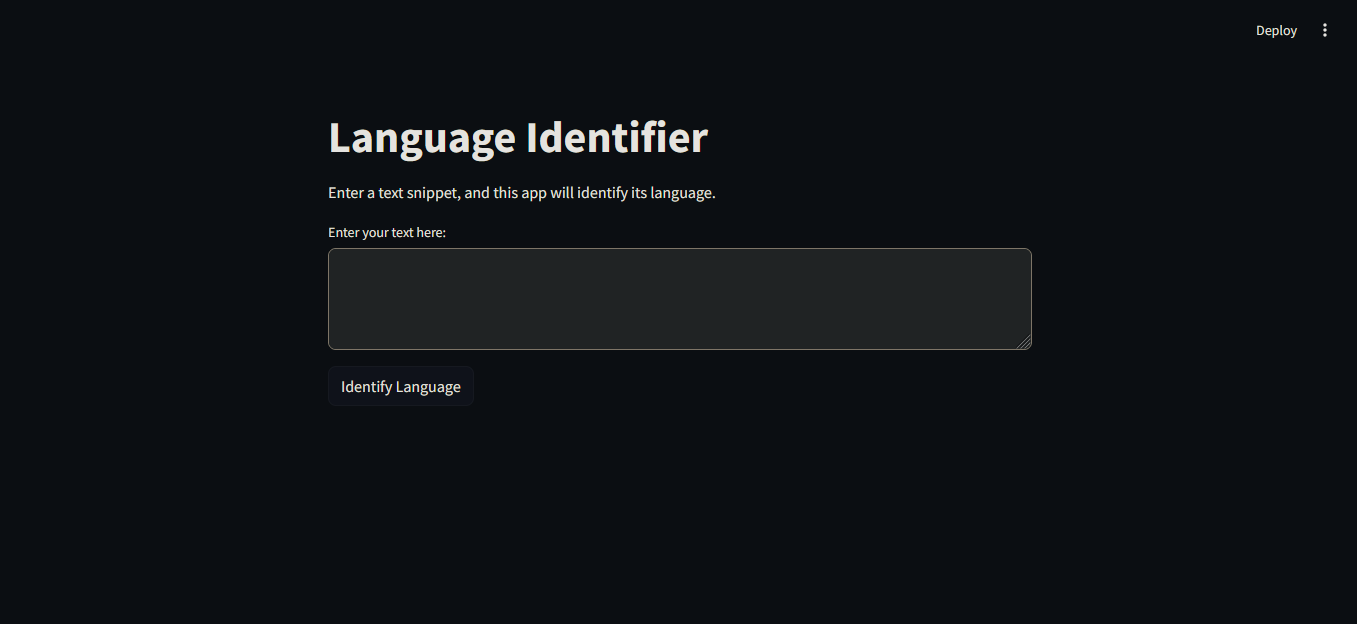
1. Karena *App* ini belum saya deploy, maka langkah pertama untuk menjalankannya adalah dengan menjalankan kode di *localhost.* Caranya dengan masuk ke Visual Studio Code > Masuk ke direktori kerja > Jalankan dengan perintah streamlit run language-identifier.py. Kenapa perintahnya seperti itu? Hal tersebut karena saya menggunakan streamlit untuk menampilkan tampilan UI dari kode yang saya buat.



Gambar 2 Running Kode

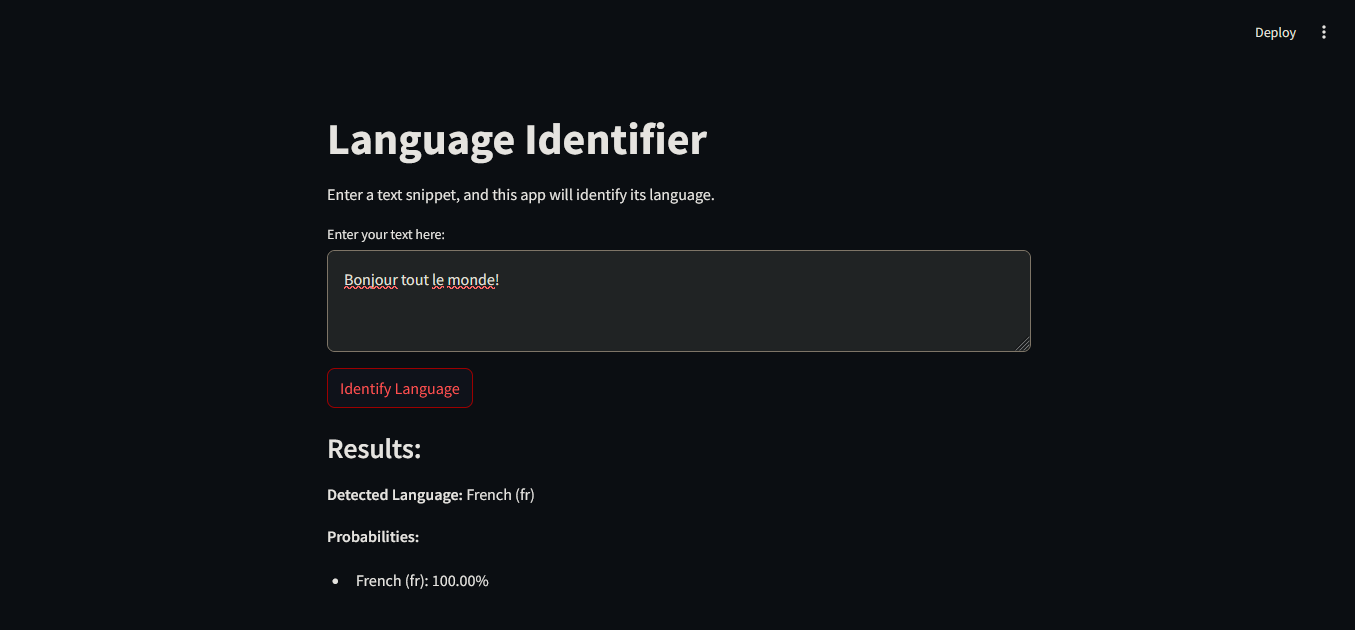
Pada gambar 2, merupakan tampilan ketika kode telah di *running.* Setelahnya, kita dapat mengakses nya pada alamat *local* yang tertera.

1. Setelah kita masuk ke halaman depan, maka akan tampil *interface* yang dapat digunakan untuk melakukan peringkasan teks, yaitu kolom *input* dan tombol proses. Seperti gambar 3 di bawah ini.



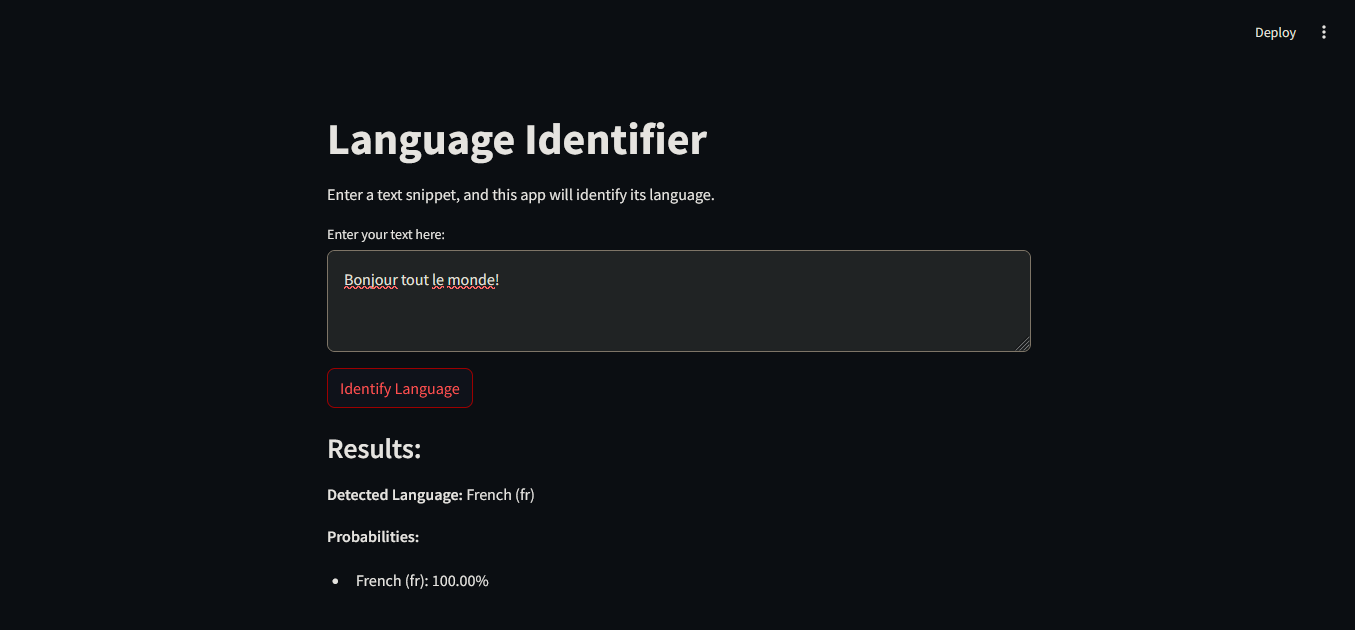
Gambar 3 Interface Language Identifier App

1. Setelahnya, pengguna dapat memasukan teks yang ingin di deteksi, conothnya seperti gambar 4 berikut ini.



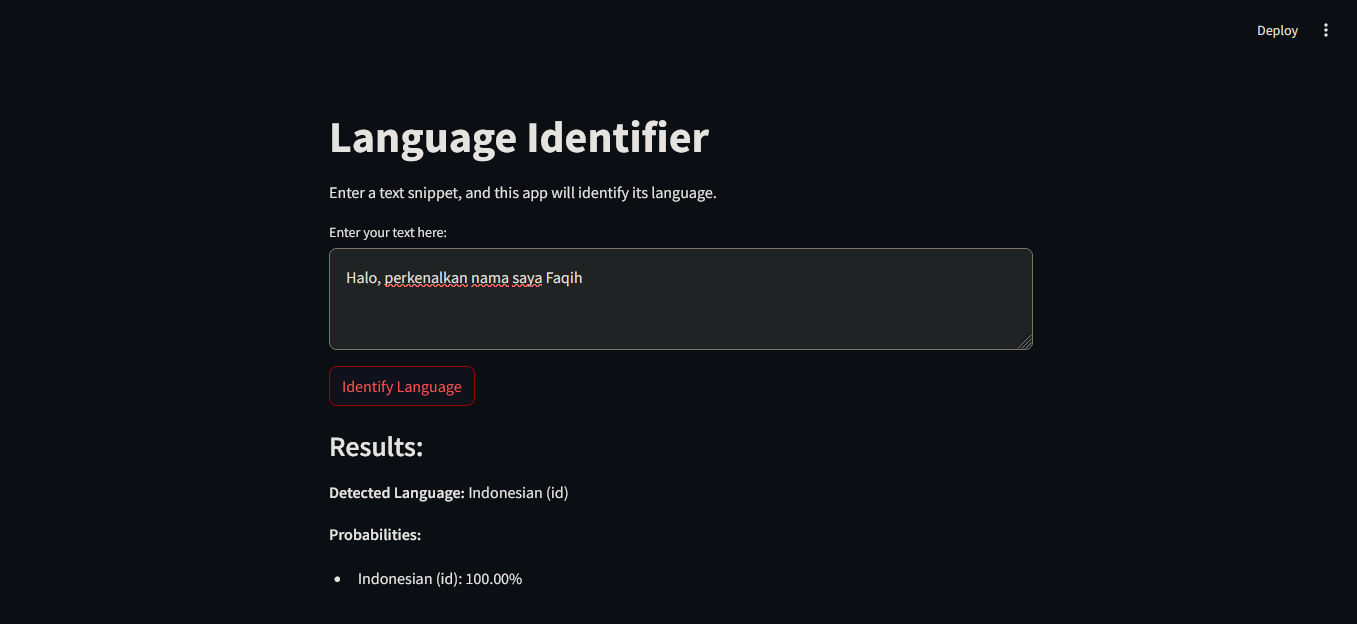
Gambar 4 Input Teks

1. Setelah pengguna memasukan teks yang akan di deteksi, pengguna dapat menekan tombol identifikasi dan sistem akan memproses teks tersebut, hingga nantinya akan ditampilkan hasil identifikasi teks seperti pada gambar 5 dibawah ini.



Gambar 5 Hasil Identifikasi

Berikut ini adalah contoh identifikasi jika bahasa Indonesia.



Gambar 6 Deteksi Teks Bahasa Indonesia