

GRID : GitHub Repository Introduction and Documentation

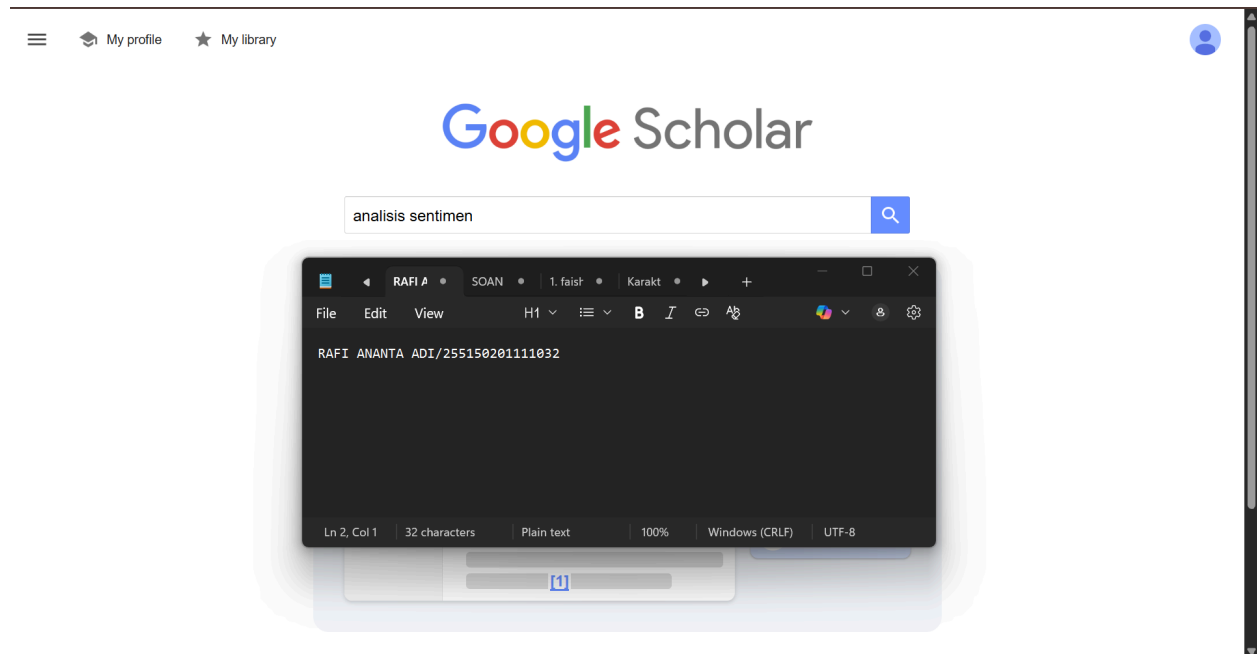
Nama : Rafi Ananta Adi

NIM : 255150201111032

Program Studi : Teknik Informatika

1. Studi Literatur dan Identifikasi Masalah

Screenshot beserta Notepad



Penjelasan

Tahap awal adalah melakukan studi pendahuluan secara mendalam. Kegiatan ini mencakup penelaahan literatur ilmiah terkait analisis sentimen, sistem rekomendasi, dan algoritma *machine learning* (seperti Naive Bayes, SVM, dan IndoBERT). Tujuannya adalah untuk mengidentifikasi dan memahami metode yang paling relevan serta efektif untuk memproses data teks ulasan dalam Bahasa Indonesia, yang seringkali memiliki karakteristik informal.

2. Pengumpulan Data (Akuisisi Data)

Screenshot beserta Notepad

Panduan dan Teknik-Teknik Web Scraping



```
File Edit View A A 8 8
RAI ANANTA ADI/255150201111032
Ln 2, Col 1 | 32 characte Plain t | 100% | Wind UTF-8
```

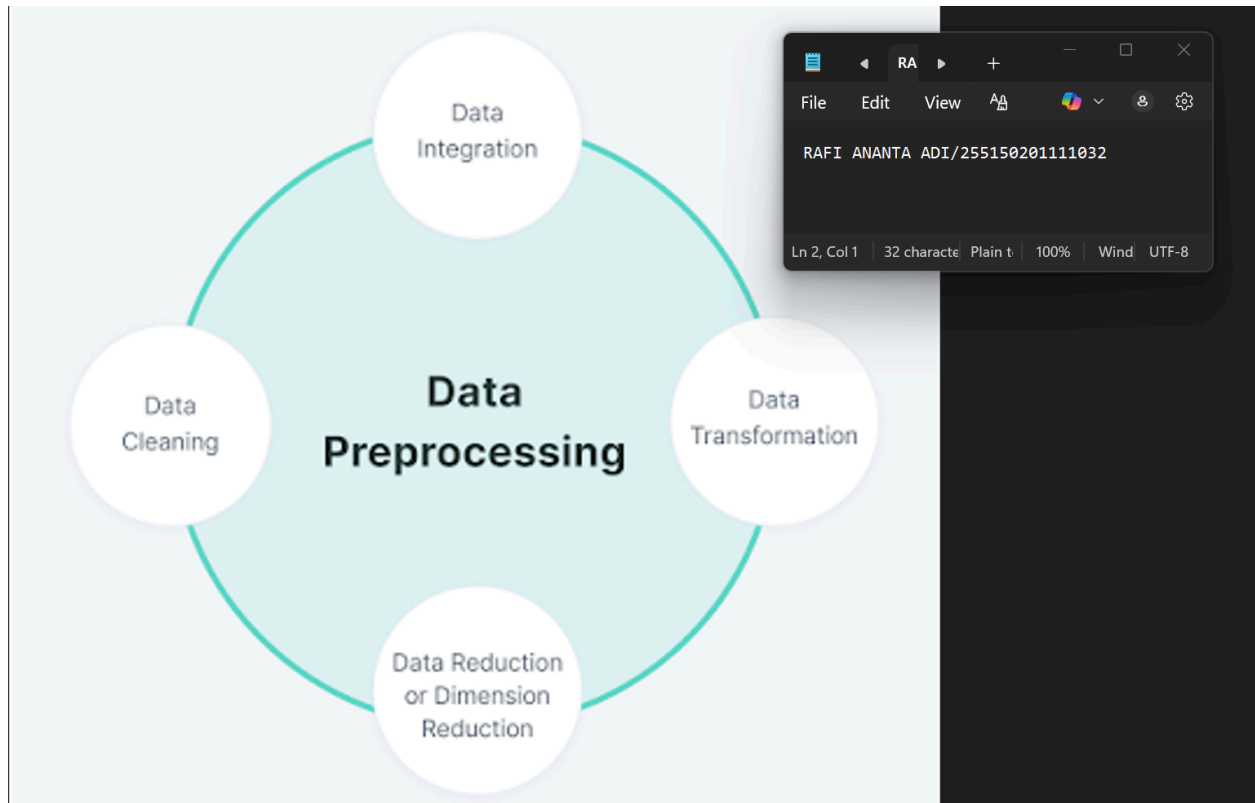
Sumber: <https://idwebhost.com/blog/panduan-teknik-web-scraping/>

Penjelasan

Langkah selanjutnya adalah akuisisi data mentah. Data ini berupa ulasan produk yang diperoleh dari platform *e-commerce* lokal (misalnya Tokopedia atau Shopee) menggunakan teknik *web scraping*. Data yang dikumpulkan mencakup komponen esensial seperti teks ulasan, rating bintang yang diberikan pengguna, nama produk, serta atribut relevan lainnya yang akan digunakan sebagai dasar analisis.

3. Pra-Pemrosesan Data (*Preprocessing*)

Screenshot beserta Notepad



Sumber: v7labs.com

Penjelasan

Data mentah yang telah dikumpulkan kemudian harus melalui tahap pembersihan dan persiapan agar siap diolah oleh model. Proses ini sangat krusial untuk menghilangkan *noise* dan menstandarisasi data.

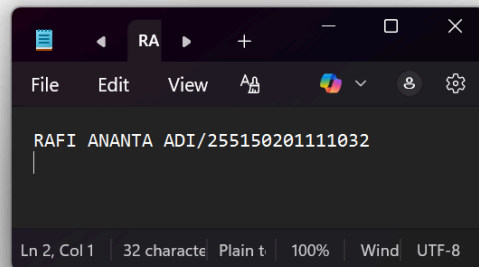
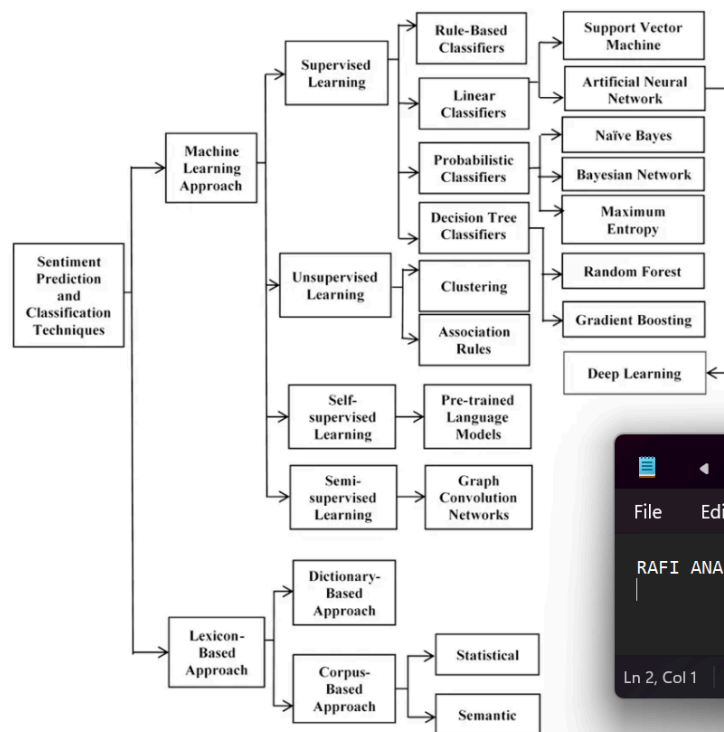
Tahapan ini meliputi:

- **Cleaning:** Eliminasi karakter khusus, angka, tautan URL, dan emoticon yang tidak relevan.
- **Case Folding:** Konversi seluruh teks menjadi format huruf kecil (lowercase) untuk uniformitas.
- **Tokenizing:** Pemecahan kalimat menjadi unit-unit kata (token) individual.
- **Stopword Removal:** Penyaringan dan pembuangan kata-kata umum yang tidak memiliki bobot makna signifikan (misalnya "dan", "yang", "di").
- **Stemming:** Transformasi kata-kata ke bentuk dasarnya (kata dasar) untuk mengurangi variasi morfologi (menggunakan pustaka seperti Sastrawi).
- **Vectorization:** Konversi data teks yang telah bersih menjadi representasi numerik (vektor) yang dapat diproses oleh algoritma *machine learning* (misalnya menggunakan metode TF-IDF).

4. Pemodelan dan Klasifikasi Sentimen

Screenshot beserta Notepad

Pendekatan yang digunakan dalam Sentiment Analysis



<https://link.springer.com/article/10.1007/s10462-023-10442-2>

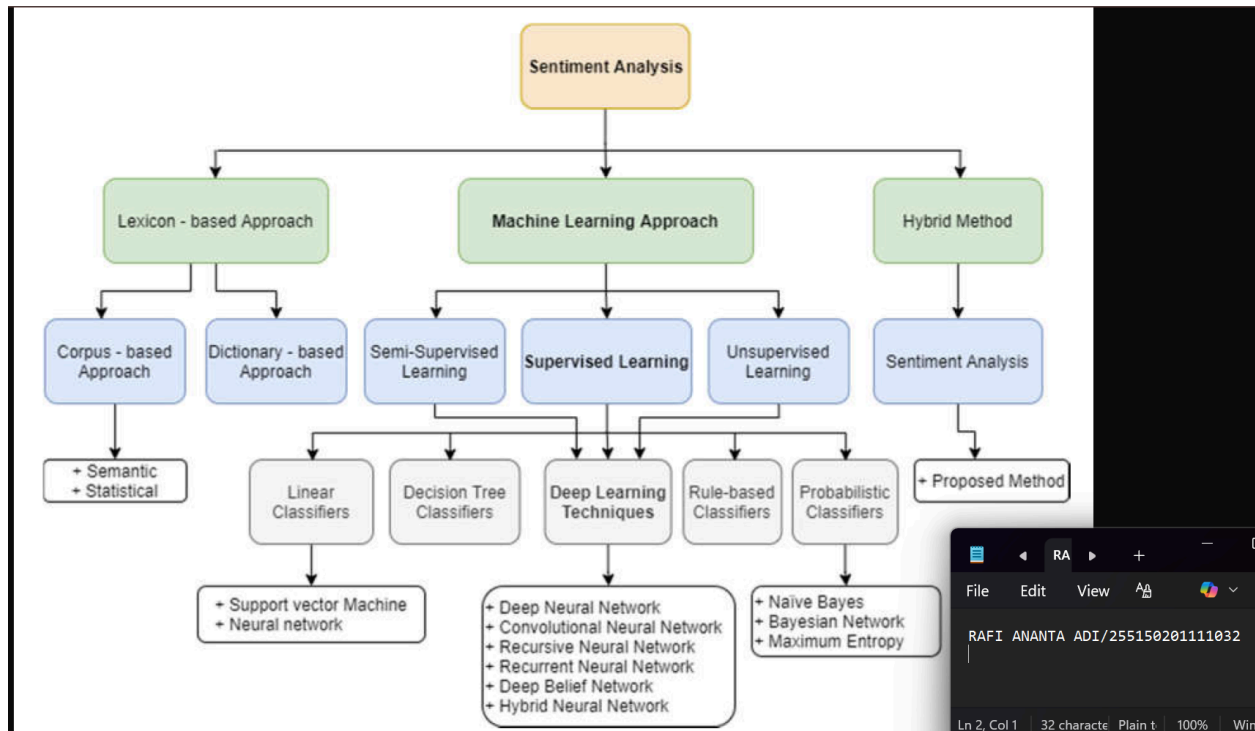
Sumber: <https://link.springer.com/article/10.1007/s10462-023-10442-2>

Penjelasan

Pada tahap ini, data numerik yang telah diproses digunakan untuk melatih model klasifikasi sentimen. Dilakukan eksperimentasi dengan membandingkan performa beberapa algoritma, mulai dari metode klasik (Naive Bayes, SVM, Logistic Regression) hingga model *deep learning* (seperti IndoBERT). Tujuannya adalah untuk menentukan model yang mampu memberikan akurasi klasifikasi tertinggi (positif, negatif, netral) terhadap data ulasan yang kompleks.

5. Integrasi dengan Sistem Rekomendasi

Screenshot beserta Notepad



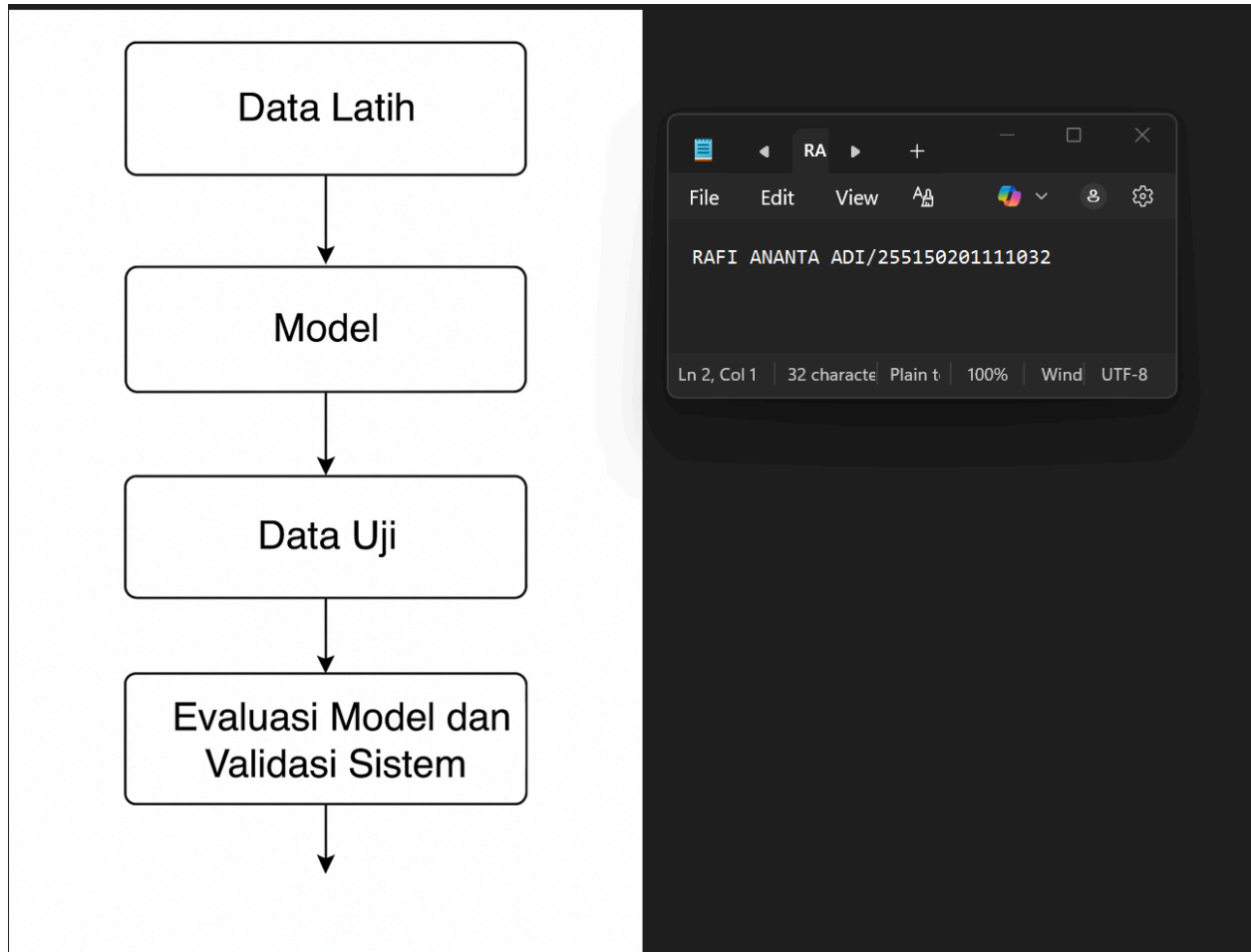
Sumber: <https://www.mdpi.com/1424-8220/21/16/5666>

Penjelasan

Hasil keluaran dari model klasifikasi sentimen (label sentimen) selanjutnya diintegrasikan ke dalam logika sistem rekomendasi produk. Integrasi ini bertujuan agar sistem tidak hanya memberikan rekomendasi berdasarkan kesamaan produk atau rating, tetapi juga secara cerdas memprioritaskan produk-produk yang memiliki riwayat sentimen ulasan positif dan sesuai dengan preferensi pengguna.

6. Evaluasi Model dan Validasi Sistem

Screenshot beserta Notepad

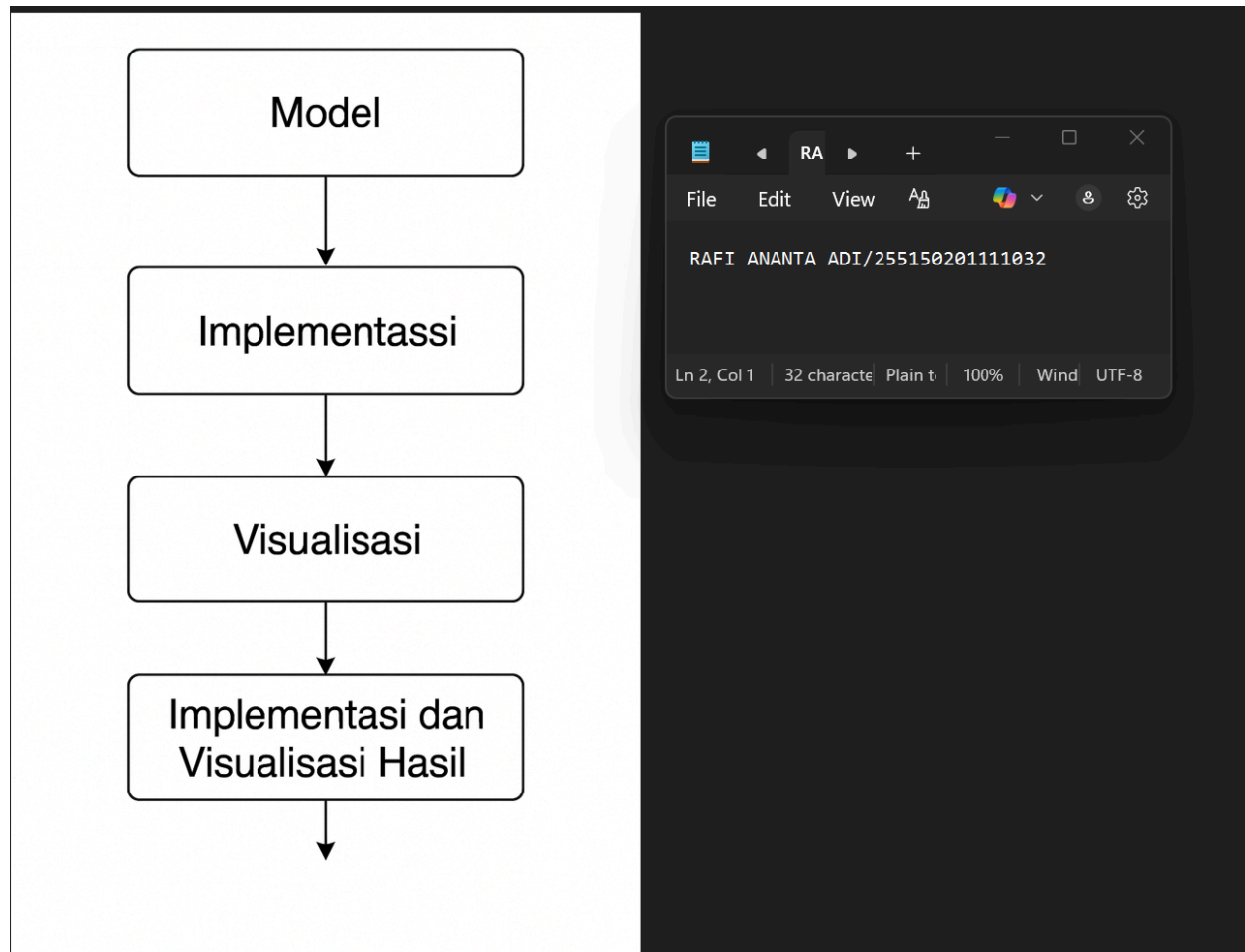


Penjelasan

Model yang telah dibangun harus diuji kinerjanya secara kuantitatif. Evaluasi dilakukan menggunakan metrik standar klasifikasi seperti *accuracy* (akurasi), *precision* (presisi), *recall*, dan *F1-score*. Selain evaluasi model, dilakukan pula validasi fungsional terhadap sistem rekomendasi untuk memastikan bahwa saran produk yang dihasilkan benar-benar relevan dan bermanfaat bagi pengguna akhir.

7. Implementasi dan Visualisasi Hasil

Screenshot beserta Notepad



Penjelasan

Langkah terakhir adalah menyajikan hasil analisis agar dapat diinterpretasi dengan mudah. Hal ini diwujudkan melalui pengembangan sebuah *dashboard* interaktif (misalnya menggunakan *framework* Streamlit atau Dash). *Dashboard* ini akan menampilkan visualisasi data seperti distribusi sentimen (misalnya dalam bentuk *pie chart*), *word cloud* untuk kata-kata yang paling sering muncul, serta daftar rekomendasi produk yang telah difilter berdasarkan sentimen.