

ANALISIS SENTIMEN

Definisi

Analisis sentimen adalah sebuah proses yang bertujuan untuk mengidentifikasi dan mengkategorikan emosi atau sentiment yang diekspresikan dalam teks tertulis. Proses ini melibatkan beberapa langkah, termasuk pemrosesan data, ekstraksi fitur, dan klasifikasi. Analisis sentiment banyak diterapkan dalam berbagai bidang, seperti ulasan produk, survei, dan komentar di media sosial, untuk memahami pandangan masyarakat terhadap suatu topik. Metode yang digunakan dalam analisis sentiment terus berkembang, dan hasilnya dapat membantu dalam pengambilan keputusan pembuatan kebijakan baru.

Tahapan

Preprocessing data adalah langkah penting dalam proses analisis sentimen, karena membantu menstandarisasi data teks dan menghilangkan elemen yang tidak relevan atau berisik. Langkah ini dapat mencakup teknik-teknik seperti stemming, lemmatization, dan penghapusan stop words dan karakter khusus. Data teks yang telah dibersihkan kemudian diubah menjadi fitur atau embedding, yang dimasukkan ke dalam pengklasifikasi untuk prediksi sentimen.

Berikut tahapan Analisis sentiment

1. Data collection

Data Crawling/web Scraping yaitu mengambil data dari media sosial (X, Instagram, Youtube, Threads, dll)

2. Preprocessing

- Data cleaning

Menghapus karakter yang tidak relevan seperti emoji, URL, mention (@user), dan simbol lainnya.

- Normalization

Mengubah kata tidak baku atau kata slang menjadi bentuk baku. Contohnya: "gk" → "tidak", "bgt" → "banget".

- Stopword Removal

Menghilangkan kata-kata umum yang tidak memberikan kontribusi besar terhadap makna kalimat, seperti “dan”, “yang”, “di”, dll

- Tokenization

Memecah teks atau kalimat menjadi unit kata-kata (token) yang dapat dianalisis oleh algoritma

- Stemming

Mengubah kata ke bentuk dasarnya. Misalnya, “berlari”, “lari-lari”, dan “pelari” akan distem menjadi “lari”.

- Translation

Menerjemahkan teks dari bahasa Indonesia ke bahasa Inggris jika menggunakan tool berbasis leksikon berbahasa Inggris seperti TextBlob. Ini berguna agar hasil analisis sentimen lebih akurat karena kamus polaritas TextBlob dibangun berdasarkan bahasa Inggris

3. Implementasi Algoritma / Metode

Menerapkan metode untuk mengklasifikasikan sentimen dari tweet yang telah diproses

4. Visualiasi Data

- Barchart
- Wordcloud

Output :

1. Presentasi (Tanya-jawab antar kelompok)
2. Laporan
3. File python lengkap beserta datasetnya (data hasil crawling, hasil cleaned, data hasil preprcesing, terjemahan ingris dan klasifikasi menggunakan metode)

Penilaian (Individu 50% , kelompok 50%)

1. Teknis 50%
2. Presentasi 20%
3. Laporan 30%