

NAMA : ADITYA ULIL ALBAB
NIM : A11.2023.15093
KELAS : A11.4703

Laporan Proyek Analisis Sentimen Ulasan Pantai Karimunjawa

1. Topik/Judul

Search Engine Berdasarkan Analisis Sentimen Ulasan Pengunjung Pantai Karimunjawa Menggunakan Model IndoRoberta

2. Latar Belakang dan Urgensi

Pariwisata merupakan sektor vital bagi ekonomi daerah, dan Pantai Karimunjawa adalah salah satu destinasi unggulan di Indonesia. Di era digital, wisatawan sering membagikan pengalaman mereka melalui ulasan daring di platform seperti Google Maps. Ulasan ini mengandung informasi berharga mengenai kepuasan pengunjung, kualitas fasilitas, dan keindahan alam.

Namun, menganalisis ribuan ulasan secara manual tidak efisien. Diperlukan metode otomatis untuk mengolah data teks tak terstruktur ini menjadi wawasan yang dapat ditindaklanjuti (actionable insights) untuk membantu pengelola wisata dan pemerintah daerah dalam meningkatkan layanan dan strategi promosi.

3. Masalah dan Objektif / Tujuan Riset

a. Masalah:

- Banyaknya data ulasan tidak terstruktur yang sulit diolah secara manual.
- Adanya variasi bahasa, singkatan, dan bahasa gaul (slang) dalam ulasan yang menyulitkan analisis standar.

b. Tujuan Riset (Objektif):

Data ulasan pengunjung Pantai Karimunjawa dikumpulkan melalui Google Maps, kemudian dilakukan pra-pemrosesan seperti pembersihan data, stemming, dan normalisasi untuk menghilangkan *noise* serta menstandarkan teks. Selanjutnya, ulasan tersebut diklasifikasikan berdasarkan sentimen menjadi kategori positif, negatif, atau netral. Proses ini bertujuan untuk menganalisis distribusi sentimen guna mendapatkan pemahaman yang lebih baik mengenai persepsi umum pengunjung terhadap Pantai Karimunjawa.

4. Dataset

a. Sumber Data: Google Maps Reviews.

- Metode Pengumpulan: Scraping menggunakan platform Apify , saya juga berikan source codennya di (Scraping.ipynb)
- Lokasi Target: Berbagai pantai di Karimunjawa (misal: Pantai Tanjung Gelam, Pantai Alano, dll).

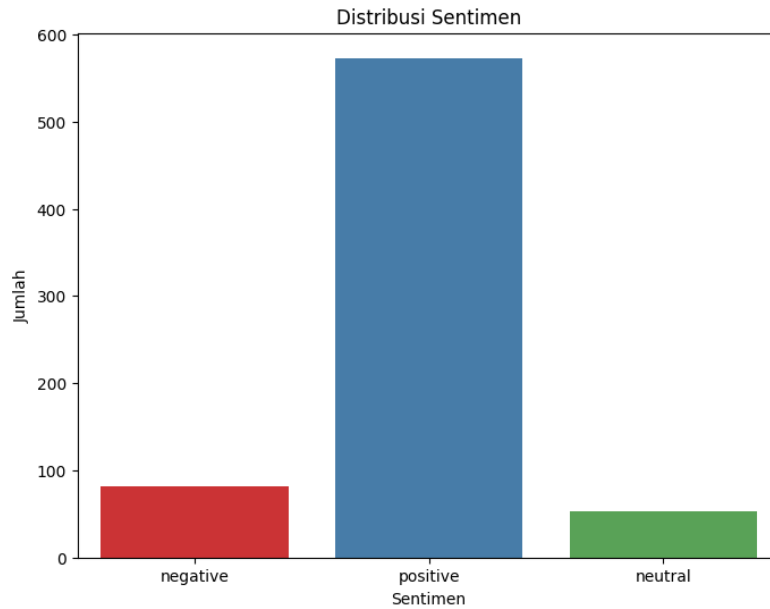
b. Alur Pemrosesan Data:

1. Raw Data: Hasil scraping mentah berisi (title, url, stars, name, review url, dan text)
2. Cleaning (Cleaning.ipynb):
 - Menghapus duplikasi dan data kosong (NaN).
 - Membersihkan karakter khusus (emoji, simbol), mention (@), hashtag, dan URL menggunakan Regular Expressions (Regex).
 - Case folding (mengubah teks menjadi huruf kecil).
3. Preprocessing (Preprocessing.ipynb):
 - Normalisasi: Mengubah kata tidak baku/singkatan (misal: "yg" -> "yang", "bagussss" -> "bagus")
 - Stopword Removal: Menghapus kata hubung yang tidak bermakna menggunakan pustaka Sastrawi.
 - Tokenized: Memisahkan kata per kata dengan koma.
 - Stemming: Mengubah kata berimbuhan menjadi kata dasar menggunakan Sastrawi.
4. Labelling (`Labelling.ipynb`)
 - Klasifikasi Sentimen: Menggunakan model IndoRoberta (w11wo/indonesian-roberta-base-sentiment-classifier) untuk mengklasifikasikan ulasan menjadi Positif, Negatif, atau Netral.

5. Hasil Analisis

Berdasarkan proses pelabelan otomatis pada dataset, didapatkan distribusi sentimen sebagai berikut:

- Positif: 584 ulasan (Dominan)
- Negatif: 82 ulasan
- Netral: 53 ulasan



Gambar 1 Grafik Distribusi Sentimen

6. Observasi

- Mayoritas pengunjung memberikan respons positif, menyoroti keindahan alam dan pengalaman yang menyenangkan.
- Jumlah ulasan negatif relatif kecil dibandingkan total ulasan, namun tetap penting untuk identifikasi area perbaikan.

7. Diskusi dan Kesimpulan

Penerapan pipeline pemrosesan data yang mencakup pembersihan menggunakan Regex, stemming Sastrawi, dan normalisasi kata terbukti efektif dalam mempersiapkan teks untuk model Transformer. Model IndoRoberta berhasil menangkap konteks sentimen dengan baik, terutama pada ulasan wisata yang sering menggunakan bahasa informal. Penggunaan kamus normalisasi (norm) sangat penting, mengingat banyaknya singkatan yang muncul dalam ulasan berbahasa Indonesia, seperti "bgt" dan "gak".

8. Kesimpulan

Proyek ini berhasil mengembangkan sistem analisis sentimen otomatis untuk pariwisata Karimunjawa. Hasilnya menunjukkan bahwa mayoritas sentimen pengunjung sangat positif. Sistem ini dapat digunakan secara terus-menerus untuk memantau tingkat kepuasan wisatawan dan mendeteksi masalah-masalah penting secara real-time melalui ulasan yang baru masuk.