

DOKUMENTASI APLIKASI SENTIMENT ANALYSIS
REVIEW HOTEL DI KOTA JAKARTA



Disusun oleh:
Yohanes Dimas Pratama
A11.2021.13254

FAKULTAS ILMU KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS DIAN NUSWANTORO

Daftar Isi

1 - Project Aplikasi.....	3
2 - Deskripsi Aplikasi	3
3 - Alasan Pembuatan Aplikasi	3
4 - Model dan Alur Tahapan yang Digunakan	4
5 - Dataset yang Digunakan.....	4
6 - Performa dari Hasil Pengujian Model	4
7 - Proses Deployment Aplikasi.....	5
8 - Kesimpulan Aplikasi	5
9 - Saran dan Rekomendasi Aplikasi.....	5

1 - Project Aplikasi

Aplikasi Sentiment Analysis: Review Hotel di Kota Jakarta

2 - Deskripsi Aplikasi

Aplikasi ini dibuat dengan tujuan khusus untuk melakukan analisis sentimen terhadap ulasan hotel di kota Jakarta. Dengan menggunakan algoritma Support Vector Classification (SVC), aplikasi ini memberikan pengguna kemampuan untuk dengan cepat dan efisien mengevaluasi umpan balik pelanggan terhadap berbagai hotel di Jakarta.

Fokus utama dari aplikasi ini adalah memberikan solusi yang otomatis dan akurat dalam menilai sentimen dari ulasan pelanggan. Algoritma SVC yang digunakan dalam model training memiliki kemampuan untuk memahami dan mengklasifikasikan sentimen positif, negatif, atau netral dengan tingkat keakuratan yang tinggi.

Pengguna aplikasi ini dapat dengan mudah mengakses informasi yang relevan tentang bagaimana pelanggan merespons berbagai aspek hotel di Jakarta, mulai dari pelayanan hingga fasilitas. Dengan demikian, pemilik hotel, manajemen, atau calon tamu dapat mengambil keputusan yang lebih baik berdasarkan analisis sentimen yang disediakan oleh aplikasi.

Kecepatan dan efisiensi dalam mengevaluasi umpan balik pelanggan menjadi nilai tambah utama dari aplikasi ini. Dengan mengotomatisasi proses analisis sentimen, pengguna dapat menghemat waktu dan upaya dalam mengumpulkan dan menganalisis ulasan secara manual.

3 - Alasan Pembuatan Aplikasi

A. Permasalahan

Dalam menghadirkan aplikasi sentiment analysis untuk ulasan hotel di Jakarta, beberapa permasalahan muncul. Pertama, fluktuasi sentimen bisa menjadi kendala utama, terutama ketika dipengaruhi oleh faktor-faktor eksternal atau tren yang bersifat sementara.

Bahasa yang beragam dan ekspresi pelanggan yang bervariasi juga menjadi tantangan dalam memastikan analisis sentimen yang akurat. Selain itu, penilaian kredibilitas ulasan merupakan perhatian penting untuk menghindari informasi yang tidak akurat atau bias.

B. Tujuan

Dengan mengatasi permasalahan sebelumnya, aplikasi sentiment analysis ini bertujuan memberikan informasi akurat dan kredibel tentang sentimen pelanggan terhadap hotel di Jakarta.

Fokusnya adalah memberikan pemahaman terperinci terhadap berbagai aspek, seperti layanan, fasilitas, dan kebersihan, sehingga pengguna dapat membuat keputusan yang lebih informasional dan sesuai dengan preferensi individu.

Secara keseluruhan, tujuan aplikasi ini adalah meningkatkan pengalaman pengguna dalam merencanakan perjalanan mereka ke Jakarta dengan menyediakan pandangan yang komprehensif dan relevan terhadap ulasan hotel di kota tersebut.

4 - Model dan Alur Tahapan yang Digunakan

A. Model Web Scrapping

Model yang berfungsi untuk mengumpulkan dataset dengan cara web scrapping.

B. Model Training SVC

Model melatih algoritma pada data ulasan, menerjemahkan teks ke representasi numerik, dan memprediksi sentimen ulasan baru.

C. Alur Tahapan SVC

1. Pemilihan dan pengumpulan data

Pilih dataset ulasan hotel dengan label sentimen (positif/negatif) dan kumpulkan data sesuai kebutuhan aplikasi.

2. Preprocessing data

Lakukan preprocessing pada data teks ulasan, termasuk penghapusan karakter khusus, konversi huruf kecil, penghapusan stopwords, dan lemmatisasi kata.

3. Pelatihan model SVC

Inisialisasi dan latih model Support Vector Classifier (SVC) menggunakan data latih, memilih kernel dan menyesuaikan hyperparameter.

4. Evaluasi model

Menggunakan model yang telah dilatih untuk mengklasifikasikan atau memprediksi kelas data baru.

5 - Dataset yang Digunakan

Dataset yang digunakan dalam pengembangan aplikasi ini diperoleh melalui ekstraksi otomatis informasi dari halaman web menggunakan metode web scraping. Untuk menyederhanakan proses ini, saya telah menggunakan Serp API, sebuah alat yang memfasilitasi ekstraksi data dari hasil pencarian mesin telusur.

Dalam hal ini, fokus pengumpulan data terpusat di daerah Jakarta dengan kategori khusus, yaitu hotel. Hal ini memastikan bahwa dataset yang dihasilkan mencakup ulasan-ulasan terkait hotel-hotel di wilayah tersebut. Adapun atribut data yang berhasil dikumpulkan melalui proses web scraping ini melibatkan informasi kunci seperti nama hotel, username pengulas, rating hotel, dan ulasan hotel.

Dengan memusatkan pengumpulan data pada lokasi tertentu dan kategori khusus, saya dapat memastikan bahwa dataset yang dihasilkan relevan dan sesuai dengan kebutuhan aplikasi analisis sentimen.

Proses ekstraksi otomatis ini memberikan dasar yang kokoh untuk melatih model analisis sentimen menggunakan algoritma SVC, sehingga aplikasi dapat memberikan evaluasi yang akurat terhadap sentimen pelanggan terhadap hotel-hotel di Jakarta. Dengan memperbarui secara berkala dataset dan menjaga kualitasnya, aplikasi dapat terus memberikan informasi yang relevan dan berguna bagi pengguna.

6 - Performa dari Hasil Pengujian Model

Hasil pengujian pada model training menggunakan algoritma Support Vector Classification (SVC) menunjukkan bahwa model ini telah mencapai tingkat akurasi, presisi, recall, dan F1-score yang

tinggi. Hasil ini menggambarkan kualitas dan kinerja yang sangat baik dari model yang telah dikembangkan.

Tingkat akurasi yang tinggi menandakan bahwa model mampu dengan tepat mengklasifikasikan sentimen dari ulasan hotel di Jakarta. Presisi yang tinggi menunjukkan bahwa ketika model mengidentifikasi sentimen tertentu, kecenderungan untuk benar-benar benar dalam klasifikasinya sangat besar. Recall yang tinggi menandakan bahwa model memiliki kemampuan untuk menangkap sebagian besar sentimen yang seharusnya diklasifikasikan sebagai positif, negatif, atau netral. F1-score yang tinggi adalah hasil kombinasi presisi dan recall, memberikan ukuran keseluruhan kualitas klasifikasi model.

7 - Proses Deployment Aplikasi

1. Instalasi Streamlit
Menginstall Streamlit ke dalam environment dengan menggunakan command -> `pip install streamlit`.
2. Buat file aplikasi Streamlit
Buat file Python (misalnya, `app.py`) yang akan menjadi script utama aplikasi Streamlit.
3. Jalankan aplikasi di local
Setelah selesai maka aplikasi akan dicoba dengan dijalankan di environment lokal dengan menggunakan perintah -> `streamlit run app.py`.
4. Deploy aplikasi ke platform cloud
Setelah aplikasi dipastikan berjalan dengan baik secara lokal, maka aplikasi dapat dideploy ke platform cloud seperti Heroku, AWS, atau Azure.

8 - Kesimpulan Aplikasi

Aplikasi Sentiment Analysis yang saya buat untuk ulasan hotel di Jakarta telah melewati pengujian dengan hasil yang memuaskan. Saya membuat 2 model disini. Model yang pertama digunakan untuk mengumpulkan dataset ulasan hotel menggunakan metode web scrapping.

Untuk model yang satunya menggunakan algoritma Support Vector Classifier (SVC), yang memberikan hasil yang baik dalam menganalisis sentimen ulasan berbahasa Inggris. Meskipun berhasil, satu kekurangan teridentifikasi, yaitu keterbatasan pada bahasa, di mana aplikasi hanya dapat menganalisis ulasan dalam bahasa Inggris.

9 - Saran dan Rekomendasi Aplikasi

Untuk meningkatkan fungsionalitas, kedepannya dapat dilakukan integrasi dengan menambahkan layanan terjemahan otomatis agar aplikasi dapat menangani ulasan dalam bahasa yang berbeda, seperti bahasa Indonesia. Dengan memperbaiki keterbatasan ini, aplikasi dapat menjadi lebih fleksibel dan dapat diakses oleh lebih banyak pengguna.