



ISIAN SUBSTANSI PROPOSAL PENELITIAN

Petunjuk: Pengusul hanya diperkenankan mengisi di tempat yang telah disediakan sesuai dengan petunjuk pengisian dan tidak diperkenankan melakukan modifikasi template atau penghapusan di setiap bagian.

JUDUL

Tuliskan judul penelitian.

Aplikasi Analisis Sentimen *Hybrid* Ulasan Produk E-Commerce Tokopedia Menggunakan Algoritma *Support Vector Machine* (SVM) Berbasis *Web Dashboard*

RINGKASAN

Isian ringkasan penelitian tidak lebih dari 300 kata yang berisi urgensi, tujuan, metode, dan luaran yang ditargetkan.

Pertumbuhan e-commerce di Indonesia, khususnya Tokopedia, menghasilkan volume ulasan produk yang masif namun tidak terstruktur. Pembeli dan penjual sering kesulitan menyimpulkan kualitas produk secara cepat karena adanya ketidakkonsistenan antara teks ulasan dan *rating* bintang (misalnya ulasan sarkas atau salah klik *rating*). Penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen otomatis yang menerapkan logika *Hybrid* (menggabungkan leksikon teks dan koreksi *rating*) serta algoritma *Support Vector Machine* (SVM) untuk klasifikasi yang lebih akurat. Metode yang digunakan meliputi pengumpulan data, *preprocessing* dengan normalisasi bahasa gaul (*slang*) spesifik *marketplace*, pembobotan kata menggunakan TF-IDF, dan klasifikasi model SVM. Luaran yang ditargetkan adalah aplikasi berbasis web (Streamlit) yang mampu menampilkan visualisasi sentimen, *wordcloud*, dan identifikasi produk terlaris secara *real-time*, serta publikasi pada jurnal nasional terakreditasi dan HKI program komputer. Aplikasi ini diharapkan dapat membantu penjual dalam evaluasi produk dan pembeli dalam pengambilan keputusan.

KATA KUNCI

Isian kata kunci maksimal 5 kata yang dipisahkan dengan tanda titik koma (;)

Analisis Sentimen; *Support Vector Machine*; Logika *Hybrid*; E-Commerce; Streamlit

PENDAHULUAN

Pendahuluan penelitian tidak lebih dari 1000 kata yang terdiri dari:

1. Latar belakang dan rumusan permasalahan yang akan diteliti
2. Pendekatan pemecahan masalah
3. *State of the art* dan kebaruan
4. Peta jalan (*road map*) penelitian 5 tahun

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver

1. LATAR BELAKANG DAN RUMUSAN MASALAH

Industri *e-commerce* di Indonesia mengalami pertumbuhan pesat, dengan Tokopedia sebagai salah satu pemimpin pasar. Ulasan pelanggan menjadi indikator krusial dalam keputusan pembelian. Namun, membaca ribuan ulasan secara manual adalah hal yang tidak efisien. Masalah utama yang sering muncul adalah: (1) Volume data teks yang besar dan tidak terstruktur (banyak singkatan/bahasa gaul), (2) Ketidaksesuaian antara rating bintang dan isi teks (bias penilaian), dan (3) Kurangnya alat visualisasi sederhana bagi UMKM untuk memantau performa produk

Rumusan masalah penelitian ini adalah:

1. Bagaimana meningkatkan akurasi analisis sentimen dengan menangani bahasa tidak baku (*slang*) dan ketidakkonsistenan antara teks dan rating?
2. Bagaimana menerapkan algoritma SVM untuk mengklasifikasikan sentimen ulasan produk *fashion* di Tokopedia?
3. Bagaimana merancang *dashboard* interaktif yang memberikan *insight* bisnis (produk terlaris, kata kunci dominan) secara otomatis?

2. PENDEKATAN PEMECAHAN MASALAH

Solusi yang ditawarkan adalah pengembangan aplikasi berbasis web menggunakan kerangka kerja *Streamlit* (Python). Pendekatan teknis meliputi:

1. **Preprocessing Lanjutan:** Penggunaan kamus normalisasi khusus untuk istilah *marketplace* (misal: "brg", "gan", "mantul") untuk menangani teks informal.
2. **Hybrid Sentiment Logic:** Algoritma pelabelan data yang tidak hanya melihat kata positif/negatif, tetapi juga memvalidasinya dengan *rating* bintang pengguna untuk mengoreksi bias (misal: teks positif tapi rating 1 dianggap netral/sarkas).
3. **Machine Learning:** Implementasi SVM dengan fitur TF-IDF untuk mempelajari pola kata dan melakukan prediksi sentimen pada data baru.

3. STATE OF THE ART DAN KEBARUAN

Penelitian terdahulu umumnya hanya menggunakan leksikon murni atau *machine learning* pada teks saja. Kebaruan (*novelty*) dari penelitian ini terletak pada **Logika Hybrid** yang mengintegrasikan validasi silang antara Teks dan Rating sebelum proses pelatihan model, serta integrasi modul **Business Intelligence** (grafik penjualan & ulasan terbanyak) dalam satu *dashboard* terpadu, berbeda dengan alat analisis sentimen konvensional yang hanya menampilkan label positif/negatif

4. PETA JALAN PENELITIAN

- **2025:** Pengembangan purwarupa aplikasi berbasis *Streamlit* untuk kategori Fashion dengan metode SVM (Fokus Penelitian Saat Ini).
- **2026:** Integrasi API Tokopedia secara *real-time* dan pengembangan Mobile App.
- **2027:** Implementasi *Deep Learning* (LSTM/BERT) untuk meningkatkan akurasi pada kalimat kompleks.

- **2028:** Perluasan domain ke seluruh kategori produk (Elektronik, Otomotif, dll) dan fitur deteksi *spam review*.
- **2029:** Komersialisasi produk sebagai layanan SaaS (*Software as a Service*) bagi UMKM Indonesia.

METODE

Isian metode atau cara untuk mencapai tujuan yang telah ditetapkan ditulis tidak melebihi 1000 kata. Bagian ini dapat dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Metode penelitian harus dibuat secara utuh dengan penahapan yang jelas, mulai dari awal bagaimana proses dan luarannya, dan indikator capaian yang ditargetkan yang tercermin dalam Rencana Anggaran Biaya (RAB).

Penelitian ini menggunakan metode kuantitatif dengan tahapan pengembangan sistem *Waterfall* yang dimodifikasi untuk *Data Science*:

1. **Pengumpulan Data:** Menggunakan dataset sekunder berupa file CSV berisi ulasan produk Tokopedia (kolom: teks, rating, nama produk, kategori, terjual).
2. **Preprocessing Data:**
 - *Cleaning:* Menghapus tanda baca, angka, dan karakter non-alfanumerik.
 - *Case Folding:* Mengubah huruf menjadi huruf kecil.
 - *Normalization:* Mengubah kata gaul (*slang*) menjadi baku menggunakan kamus khusus (dictionary-based).
 - *Stopword Removal:* Menghapus kata umum yang tidak bermakna sentimen.
3. **Pelabelan Data (Hybrid Approach):** Pemberian label awal (Positif, Negatif, Netral) menggunakan aturan logika:
 - Jika sentimen leksikon = Positif DAN Rating ≤ 3 , maka Label = Netral.
 - Jika sentimen leksikon = Negatif DAN Rating ≥ 4 , maka Label = Netral.
 - Selain itu mengikuti skor leksikon.
4. **Ekstraksi Fitur & Pemodelan:**
 - Menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk pembobotan kata.
 - Melatih model SVM (*Support Vector Machine*) kernel Linear dengan pembagian data latih dan uji (80:20).
5. **Evaluasi & Visualisasi:**
 - Mengukur performa model menggunakan *Confusion Matrix* dan *Classification Report* (Akurasi, Presisi, Recall).
 - Membangun antarmuka pengguna (GUI) dengan *Streamlit* untuk menampilkan grafik, *wordcloud*, dan tabel interaktif.

JADWAL PENELITIAN

Jadwal penelitian disusun berdasarkan pelaksanaan penelitian, ditulis dengan mengisi langsung tabel berikut dan diperbolehkan menambahkan baris sesuai banyaknya jenis kegiatan.

No.	Nama Kegiatan	Bulan					
		1	2	3	4	5	6
1.	Studi Literatur & Pengumpulan Dataset						
2.	Preprocessing & Pelabelan Data						
3.	Pengembangan Model SVM						
4.	Pembuatan Dashboard Streamlit						
5.	Pengujian & Evaluasi Sistem						
6.	Penyusunan Laporan & Publikasi						

LUARAN DAN TARGET CAPAIAN

Tuliskan target luaran wajib dan tambahan (jika ada) yang akan dihasilkan.

No.	Kategori Luaran	Jenis Luaran	Target Capaian
1.	Artikel Ilmiah	Jurnal Nasional Terakreditasi (SINTA)	Submitted/Accepted
2.	HKI	Hak Cipta Program Komputer	Granted (Bersertifikat)
3.	Produk Teknologi	Prototipe Aplikasi Web	Tersedia (Localhost/Hosted)

RENCANA ANGGARAN BIAYA

Total RAB: Rp 10.000.000

Jenis Pembelajaan	Komponen	Item	Kuantitas	Biaya Satuan	Total
Belanja Bahan	ATK	Kertas, Tinta, Jilid Laporan Akhir	1 Paket	750.000	750.000
Belanja Bahan	Data & Referensi	Kuota Internet High-Speed & Akses Jurnal Internasional	1 Paket	2.000.000	2.000.000

Sewa Peralatan	Sewa Server/Hosting	Cloud Hosting (VPS) & Domain .id (1 Tahun)	1 Paket	1.250.000	1.250.000
Analisis Data	Honor Pengolah Data	Jasa Programmer (Python & Streamlit Expert)	1 Paket	3.000.000	3.000.000
Pelaporan Penelitian	Biaya Publikasi	Biaya APC (Article Processing Charge) Jurnal SINTA 2	1 Paket	2.500.000	2.500.000
Lainnya	Pendaftaran HKI	Pendaftaran Hak Cipta Program Komputer	1 Sertifikat	500.000	500.000

DAFTAR PUSTAKA

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver. Sumber pustaka mengutamakan hasil penelitian pada jurnal ilmiah yang terkini (maksimal 5 tahun terakhir). Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka.

1. Liu, B. (2020). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
2. Turban, E., et al. (2018). *Electronic Commerce 2018: A Managerial and Social Networks Perspective*. Springer.
3. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. (Referensi TF-IDF & SVM)
4. Documentation Streamlit. (2024). *Streamlit: The fastest way to build and share data apps*. <https://streamlit.io/>
5. Syahputra, H., & Buanawati, A. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Tokopedia Menggunakan Algoritma Support Vector Machine. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1), 12-18
6. Wibowo, A., & Khairina, D. M. (2021). Analisis Sentimen Ulasan Produk Fashion pada E-Marketplace Tokopedia Menggunakan Metode Support Vector Machine (SVM). *Jurnal Sains dan Teknologi Komputer*, 12(2), 88-96.

