

Business Requirement Document (BRD): Pendekatan Natural Language Processing Dengan SVM Dan Naïve Bayes Untuk Deteksi Berita Palsu Bahasa Indonesia.

1. Executive Summary

Sistem ini dirancang untuk mengembangkan prototype **deteksi berita palsu bahasa Indonesia** dengan pendekatan **NLP** dan **ML**, dengan fokus utama pada **perbandingan kinerja algoritma Naive Bayes dan Support Vector Machine (SVM)**. Kedua algoritma dipilih karena kesederhanaan, keefektifan, dan kemampuannya menangani karakteristik teks berbahasa Indonesia.

2. Latar Belakang dan Justifikasi Bisnis

- **Kesenjangan Algoritma:** Meskipun banyak algoritma tersedia, belum ada kajian mendalam yang membandingkan Naive Bayes dan SVM dalam konteks deteksi hoaks bahasa Indonesia.
- **Karakteristik Bahasa Indonesia:** Kedua algoritma akan dievaluasi berdasarkan kemampuan menangani fitur-fitur linguistik khas Indonesia seperti prefiks, konfiks, dan kosakata informal.

3. Tujuan dan Sasaran Bisnis (SMART)

Sasaran Spesifik:

1. **Specific:** Membandingkan performa algoritma **Naive Bayes** dan **SVM** menggunakan ekstraksi fitur NLP (TF-IDF, Word Embeddings).
2. **Measurable:** Metrik evaluasi: **Akurasi, Presisi, Recall, F1-Score, ROC-AUC**.
3. **Achievable:** Target akurasi $\geq 85\%$, F1-score $\geq 85\%$ pada dataset uji.
4. **Relevant:** Menghasilkan rekomendasi algoritma terbaik untuk klasifikasi teks berita Indonesia.
5. **Time-bound:** Eksperimen dan evaluasi selesai dalam 4 bulan pertama.

4. Scope (Ruang Lingkup)

Dalam Cakupan:

- Eksperimen dengan **Naive Bayes** dan **SVM**.
- Ekstraksi fitur: **TF-IDF, Word Embeddings (Word2Vec/FastText)**.
- Analisis stylometric (fitur statistik teks: panjang kalimat, frekuensi kata).
- Evaluasi model dengan metrik yang relevan.

Di Luar Cakupan:

- Algoritma lain seperti Neural Network atau Logistic Regression.
- Analisis gambar/video.
- Deployment skala produksi.

5. Functional Requirements (Kebutuhan Fungsional)

1. Manajemen Data & Preprocessing NLP

- FR-01: Impor dataset berita hoaks & valid.
- FR-02: Pra-pemrosesan teks: normalisasi, tokenisasi, stemming, stopword removal.

2. Ekstraksi Fitur & Pemodelan

- FR-03: Ekstraksi fitur menggunakan TF-IDF dan Word Embeddings.
- FR-04: Pelatihan model Naive Bayes dan SVM.
- FR-05: Evaluasi model dengan metrik: Akurasi, F1-Score, ROC-AUC.

3. Prediksi & Interpretasi

- FR-06: Klasifikasi teks baru dengan confidence score.
- FR-07: Menampilkan kata kunci yang memengaruhi keputusan model (feature importance).

6. Non-Functional Requirements

1. **Performance:** Waktu respons inferensi ≤ 2 detik.
2. **Interpretability:** Menyediakan penjelasan sederhana mengapa teks diklasifikasikan sebagai hoaks/valid.
3. **Scalability:** Kode modular untuk memungkinkan penambahan algoritma di masa depan.
4. **Reproducibility:** Eksperimen harus dapat diulang dengan hasil konsisten.

7. Kesimpulan

Dokumen ini mengusulkan pengembangan sistem deteksi berita palsu dengan fokus pada **perbandingan algoritma Naive Bayes dan SVM** menggunakan pendekatan NLP. Hasilnya diharapkan dapat menjadi panduan dalam pemilihan algoritma yang efektif untuk tugas klasifikasi teks berbahasa Indonesia, serta kontribusi bagi riset NLP di Indonesia.