

**"Implementasi Sistem Temu Kembali Informasi Menggunakan  
Metode Klasifikasi Teks untuk Menyaring Berita Berdasarkan  
Kategori"**



Daniel Nomolas Wicaksono(A11.2022.14448)

Program Studi Pembelajaran Mesin FIK Universitas Dian Nuswantoro

Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50131

**Deskripsi Singkat:**

Tugas ini berfokus pada pengembangan sistem temu kembali informasi berbasis klasifikasi teks yang mampu menyaring berita daring dalam berbagai kategori, seperti politik, ekonomi, olahraga, teknologi, dan hiburan. Sistem ini bertujuan untuk mengelompokkan berita sesuai kategori yang diinginkan, memudahkan pengguna menemukan informasi yang relevan dengan cepat. Penelitian ini mencakup pengumpulan dan pra-pemrosesan data teks, seleksi fitur, pemilihan dan evaluasi model klasifikasi, serta analisis hasil untuk menentukan pendekatan terbaik dalam sistem temu kembali informasi.

**Masalah dan Tujuan:**

Masalah utama yang ingin diselesaikan adalah kesulitan dalam menyaring informasi yang relevan di tengah banyaknya berita daring yang terus diperbarui. Pengguna seringkali dihadapkan dengan tumpukan informasi yang mungkin tidak relevan atau sulit diakses berdasarkan kategori yang diminati. Tujuan dari penelitian ini adalah untuk membangun sistem klasifikasi teks yang dapat meningkatkan akurasi dan relevansi pencarian berita daring sesuai kategori sehingga pengguna lebih mudah menemukan informasi sesuai kebutuhan mereka.

**Penjelasan Dataset:**

Dataset yang digunakan adalah dataset berita daring yang telah diklasifikasikan berdasarkan kategori, seperti **BBC News**, **AG News Dataset**, atau dataset lain yang berisi berbagai artikel berita. Fitur yang digunakan meliputi:

- **Judul Berita:** Memberikan konteks singkat untuk membantu klasifikasi kategori.
- **Isi Berita (Teks Utama):** Menyediakan informasi inti untuk menentukan relevansi dan akurasi kategori.
- **Kategori (Label):** Kategori target seperti politik, ekonomi, olahraga, hiburan, teknologi, dan lain-lain.

Setiap fitur akan dianalisis untuk mengoptimalkan hasil temu kembali informasi.

### Metode Eksperimen:

Penelitian ini menggunakan metode pembelajaran mesin untuk klasifikasi teks dengan tahapan sebagai berikut:

1. **Pengumpulan Data:** Dataset dikumpulkan dari sumber-sumber daring atau dataset berita yang telah dipublikasikan secara terbuka. Proses ini juga mencakup pembersihan data untuk menghapus data yang tidak relevan atau redundan.
2. **Pra-Pemrosesan Data Teks:** Dilakukan serangkaian proses seperti tokenisasi, normalisasi teks (lowercasing, stemming, dan lemmatization), penghapusan tanda baca dan stop words. Tahapan ini penting untuk meningkatkan kualitas representasi data dalam model.
3. **Ekstraksi dan Seleksi Fitur:** Data teks diubah menjadi representasi vektor menggunakan metode **TF-IDF (Term Frequency - Inverse Document Frequency)** atau **word embeddings** seperti Word2Vec atau GloVe. Representasi ini membantu model memahami konteks kata dan hubungannya.
4. **Pemilihan dan Pelatihan Model:** Model pembelajaran mesin seperti **Naive Bayes**, **Support Vector Machine (SVM)**, **k-Nearest Neighbors (kNN)**, atau **Deep Learning** (seperti LSTM atau CNN) diterapkan. Setiap model diuji untuk melihat model mana yang memberikan hasil terbaik dalam klasifikasi berita.
5. **Evaluasi Kinerja Model:** Model yang dilatih akan dievaluasi menggunakan metrik seperti **akurasi**, **presisi**, **recall**, dan **F1-score**. Evaluasi ini bertujuan menilai kinerja model dalam mengklasifikasikan berita dengan benar ke dalam kategori yang sesuai.
6. **Optimisasi Model:** Model akan dioptimalkan menggunakan **tuning hyperparameters** untuk meningkatkan akurasi dan kinerja keseluruhan. Proses optimisasi dapat mencakup teknik seperti cross-validation dan grid search.
7. **Pengujian Sistem Temu Kembali Informasi:** Sistem diuji dengan input pengguna berupa kata kunci atau kategori untuk melihat apakah hasil yang diberikan relevan dan sesuai kebutuhan pencarian.
8. **Analisis Hasil dan Kesimpulan:** Analisis hasil dilakukan untuk mengidentifikasi faktor-faktor yang mempengaruhi akurasi dan relevansi, seperti pemilihan fitur dan algoritma. Penelitian ini diakhiri dengan kesimpulan terkait efektivitas pendekatan yang digunakan dan saran untuk pengembangan lebih lanjut.

### Ekspektasi Output:

Output yang diharapkan adalah sistem temu kembali informasi yang dapat mengklasifikasikan berita sesuai kategori dengan akurasi yang tinggi. Sistem ini juga harus mampu memberikan hasil pencarian berita yang relevan sesuai dengan kategori atau kata kunci yang diinginkan oleh pengguna.