

Framework NLP untuk Analisis Topik dan Tren dalam Artikel Teknologi



Oleh :

Andika Wira Yumna (21081010002)

Dosen Pengampu :

Prof. Dr. Basuki Rahmat, S.Si, MT, ITS-AI

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2024

ABSTRAK

Seiring dengan perkembangan teknologi dan volume artikel teknologi yang terus meningkat, pengolahan dan analisis teks secara otomatis menjadi kebutuhan penting untuk mendapatkan wawasan berharga. Proposal ini menyajikan pengembangan framework berbasis Natural Language Processing (NLP) untuk mengidentifikasi topik utama dan tren dalam artikel teknologi. Framework yang diusulkan menggabungkan teknik-teknik NLP modern, seperti tokenisasi, penghapusan stopwords, dan algoritma berbasis model transformer seperti BERT, untuk analisis topik. Dengan penerapan ini, diharapkan dapat memberikan kontribusi signifikan bagi peneliti maupun praktisi teknologi.

Kata Kunci: Natural Language Processing, Analisis Topik, Tren Teknologi, Framework NLP, Artikel Teknologi.

BAB I: PENDAHULUAN

Latar Belakang

Dalam era digital saat ini, artikel teknologi menjadi sumber informasi yang sangat penting bagi berbagai pihak, mulai dari peneliti hingga pelaku bisnis. Namun, meningkatnya jumlah artikel yang dipublikasikan setiap hari membuat analisis manual menjadi tidak efisien. Dengan adanya metode NLP, proses analisis dapat dilakukan secara otomatis untuk menyaring informasi yang relevan dan mendalam.

Rumusan Masalah

1. Bagaimana cara mengidentifikasi topik utama dalam artikel teknologi secara otomatis?
2. Bagaimana memvisualisasikan tren teknologi berdasarkan artikel yang dianalisis?

Tujuan Penelitian

1. Mengembangkan framework NLP untuk menganalisis artikel teknologi.
2. Mengidentifikasi topik utama dan tren yang berkembang dalam artikel teknologi.

Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Membantu peneliti untuk memahami perkembangan teknologi secara efisien.
2. Menyediakan alat bantu analisis otomatis untuk mempermudah pengambilan keputusan berdasarkan data tekstual.

BAB II: TINJAUAN PUSTAKA

Natural Language Processing (NLP)

NLP adalah cabang ilmu komputer yang berfokus pada interaksi antara komputer dan bahasa manusia. Teknik dasar seperti tokenisasi, stopwords removal, dan stemming memainkan peran penting dalam preprocessing teks.

Model Berbasis Transformer

Model seperti BERT dan DistilBERT memungkinkan pemahaman konteks dua arah dalam analisis teks, sehingga meningkatkan akurasi analisis dibandingkan dengan pendekatan tradisional.

Analisis Topik

Teknik seperti Latent Dirichlet Allocation (LDA) dan clustering berbasis K-means digunakan untuk mendeteksi pola dalam kumpulan artikel. Analisis temporal sering diterapkan untuk mengamati perubahan distribusi topik dari waktu ke waktu.

Tren Teknologi

Analisis tren teknologi dilakukan dengan melacak prevalensi topik tertentu berdasarkan data historis. Hal ini penting untuk memahami arah perkembangan teknologi.

BAB III: METODE PENELITIAN

Pengumpulan Data

Artikel teknologi dikumpulkan menggunakan web scraping dari sumber seperti TechCrunch, Wired, dan Medium. Data yang terkumpul mencakup artikel dalam rentang waktu 5 tahun terakhir.

Preprocessing Data

1. **Tokenisasi:** Memecah teks menjadi unit kata.
2. **Stopword Removal:** Menghapus kata-kata umum yang tidak relevan.
3. **Embedding:** Menggunakan model BERT untuk merepresentasikan teks dalam format vektor yang dapat dianalisis.

Model Analisis Topik

Framework menggunakan dua metode utama:

1. **LDA:** Untuk mengidentifikasi topik tersembunyi dalam artikel.

2. **K-means Clustering:** Untuk mengelompokkan artikel berdasarkan kesamaan tema.

Analisis Tren

Distribusi topik dianalisis menggunakan metode temporal untuk mengidentifikasi perubahan minat terhadap topik tertentu sepanjang waktu.

Evaluasi Model

Evaluasi dilakukan dengan membandingkan hasil analisis model dengan data manual. Metode evaluasi meliputi:

1. **Koherensi Topik:** Mengukur relevansi antar kata dalam topik.
2. **Akurasi Model:** Menggunakan data anotasi manual sebagai pembanding.