

Review Artikel Ilmiah

Nama : Putu Bagus Arjawa

NIM : 2308561058

Kelas : D

1. Informasi Artikel

- Judul Artikel:

Comparative Analysis of Machine Learning Models for Fake News Detection in Social Media

- Penulis:

Bahaa Eddine Elbaghazaoui, Mohamed Amnai, Youssef Fakhri, Ali Choukri², Noreddine Gherabi

- Jurnal:

IAES International Journal of Artificial Intelligence (IJ-AI)

Vol. 14, No. 3, June 2025, pp. 1951~1959

ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i3.pp1951-1959

<https://ijai.iaescore.com/index.php/IJAI/article/view/26867/1453>

2. Latar Belakang (Background)

(Tuliskan latar belakang dari artikel tersebut secara ringkas, menjelaskan konteks penelitian, masalah yang diteliti, dan relevansi topiknya. Gunakan bahasa Anda sendiri tanpa copy-paste langsung dari artikel.)

Sosial media saat ini sudah menjadi bagian yang penting dalam kehidupan manusia, tidak hanya berperan sebagai media komunikasi, tetapi juga sebagai jembatan untuk mendapatkan informasi. Seperti pisau, bilahnya bermata dua, sisi baik dan buruknya. Di sisi positif manusia dapat mencari informasi dengan mudah di media sosial, buruknya oknum tertentu menyebarkan berita bohong dan mengada-ada. Itu sebabnya penelitian dilakukan, dengan harapan bisa membuat sistem deteksi berita bohong yang efektif dan skalabel dengan menggunakan SVC Support Vector Classifier linear dengan dataset sosial media yang besar. Hasil penelitian ini berupa performa, akurasi dan waktu eksekusi. Potensi dari penelitian ini adalah menghasilkan sistem yang real-time dalam mengklasifikasikan berita bohong dan diimplementasikannya ke sistem yang lebih besar di masa depan.

3. Rumusan Masalah

Seberapa besar hasil akurasi yang diperoleh oleh masing-masing model pembelajaran mesin yang diuji?

Seberapa Besar waktu eksekusi yang dihasilkan oleh masing-masing model pembelajaran mesin yang diuji?

4. Batasan Masalah (Jika Ada)

Dataset ini terdiri dari total 44.898 artikel berita, dengan 21.417 di antaranya adalah berita asli yang bersumber dari platform kredibel seperti Reuters. Sisa 23.481 artikel merupakan berita palsu yang dikumpulkan dari situs web tidak tepercaya, sebagaimana diidentifikasi oleh Politifact dan Wikipedia. Seluruh artikel dalam dataset ini terutama diterbitkan antara tahun 2016 dan 2017, mencakup periode peristiwa politik penting. Data disajikan dalam format CSV, mencakup kolom untuk Title (judul), Text (isi artikel lengkap), dan Label biner yang menunjukkan apakah artikel tersebut asli (1) atau palsu (0).

5. Metodologi Penelitian (Research Methodology)

5.1 Jenis Penelitian

Penelitian yang dilakukan merupakan *applied research*, dengan pendekatan kuantitatif eksperimental. Fokus utamanya adalah membangun dan mengevaluasi masing-masing model pembelajaran mesin yang dipilih untuk diujikan akurasi dan waktu eksekusinya.

5.2 Metode Pengumpulan Data

Data dikumpulkan di ISOT Fake News Dataset yang sumbernya berasal dari kaggle.

5.3 Langkah-langkah

5.3.1 Data Cleaning

Data akan dilakukan proses cleaning data, sebelum dilatih ke model. Prosesnya adalah text processing, tokenization, stopword removal, dan stemming dan lemmenization.

5.3.2 Pelatihan dan Evaluasi

Data kemudian dilatih ke model ML yang diujikan. Metrik evaluasi yang dilakukan di dalam penelitian penulis adalah accuracy, recall, f-1 score, dan durasi.

5.3 Alat dan Perangkat yang Digunakan

Bahasa yang digunakan dalam pelatihan adalah Python.

6. Desain Eksperimen / Uji Coba (Experimental Design / Testing Setup)

Desain pada eksperimen tersebut berfokus pada evaluasi kuantitatif kinerja berbagai model *machine learning* untuk deteksi hoaks. Dataset berlabel akan dibagi menjadi set pelatihan (80%) dan set pengujian (20%), dengan representasi kelas yang seimbang. Parameter yang diukur meliputi akurasi, presisi, recall, F1-score (dalam persentase), serta durasi eksekusi setiap model (dalam detik). Parameter model akan diatur ke konfigurasi standar atau yang optimal.

7. Hasil Eksperimen dan Pembahasan (Results and Discussion)

Dari hasil eksperimen yang dilakukan oleh penulis, ditemukan bahwa beberapa model mencapai akurasi prediksi di atas 99% untuk deteksi hoaks. Meskipun Gradient Boosting menunjukkan akurasi dan F1-score tertinggi (99.61% dan 99.60%), durasi eksekusinya sangat lambat (159.8904 detik). Sebaliknya, Linear SVC mencapai performa yang sangat kompetitif (akurasi 99.43%, F1-score 99.42%) dengan waktu eksekusi yang jauh lebih efisien, yaitu hanya 21.2639 detik. Mengingat kebutuhan akan keseimbangan antara akurasi tinggi dan efisiensi komputasi untuk deteksi berita hoaks dan kategori wilayah, Linear SVC dipilih sebagai metode utama dalam proposal ini, menawarkan kombinasi yang optimal dalam keandalan dan kecepatan.

8. Kesimpulan (Conclusion)

Tentu, ini ringkasan kesimpulan dalam Bahasa Indonesia untuk review tugas Anda:

Ringkasan Kesimpulan

Penelitian ini berhasil mengatasi empat tantangan big data (pengumpulan, penyimpanan, analisis, dan prediksi) dengan fokus pada deteksi berita palsu berbasis teks di media sosial. Berbagai model machine learning diuji, dan Gradient Boosting terbukti paling efektif dengan akurasi 99,61%. Model XGBoost juga menunjukkan kinerja tinggi (99,59%) dan keseimbangan baik antara akurasi dan waktu eksekusi, menjadikannya pilihan ideal untuk aplikasi yang membutuhkan presisi dan kecepatan. Konsistensi kedua model ini menunjukkan ketangguhan mereka dalam memproses data kompleks.

Namun, penelitian ini memiliki keterbatasan karena hanya berfokus pada teks dan belum mencakup elemen multimedia seperti gambar atau video yang sering terlibat dalam misinformasi. Dataset yang digunakan juga terbatas pada periode dan domain tertentu. Untuk masa depan, disarankan untuk mengintegrasikan analisis gambar dan video, serta mengembangkan sistem deteksi berita palsu secara real-time untuk solusi yang lebih dinamis.