

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Dalam beberapa tahun terakhir, *buzzer* di media sosial, terutama di *TikTok*, memainkan peran penting dalam memperkuat pesan yang di sampaikan dalam sebuah postingan, baik untuk kepentingan komersial maupun politik. *Buzzer* ini bisa berupa individu atau akun yang secara aktif mempromosikan produk atau agenda tertentu. Siapa pun yang memiliki pengikut signifikan dapat berperan sebagai *buzzer*, yang membuat aktivitas mereka semakin sulit diidentifikasi[1].

*Buzzer* sering menggunakan akun palsu untuk menciptakan ilusi dukungan publik, yang dapat memengaruhi opini masyarakat secara luas. Aktivitas ini berpotensi memicu polarisasi, terutama dalam konteks politik. Oleh karena itu, identifikasi *buzzer* penting untuk mencegah penyebaran disinformasi dan menjaga kestabilan informasi di media sosial[2].

Beberapa algoritma seperti *Random Forest (RF)*, *SVM*, *Decision Tree (DT)*, dan *Logistic Regression (LR)* digunakan untuk modeling. *RF* terbukti memberikan akurasi terbaik di atas 90% dalam kasus klasifikasi, sementara algoritma lainnya berkisar antara 70%-80% [3]. Evaluasi model dilakukan dengan metrik seperti akurasi, presisi, recall, dan *F1-score*, serta menggunakan validasi silang atau *cross-validation* untuk memastikan model tidak *overfitting* [4].

Dalam penelitian ini, metode *Cosine Similarity* digunakan untuk mengukur kemiripan antar-komentar dalam sebuah data postingan di *TikTok* [5]. Algoritma *K-means* kemudian digunakan untuk mengelompokkan komentar yang memiliki pola serupa, membantu mendeteksi *buzzer* dengan lebih akurat [6]. Teknik ini memanfaatkan berbagai fitur seperti nilai yang di hasilkan *Cosine Similarity*, jumlah komentar serupa, total komentar yang diposting, serta selisih waktu antar-komentar dari setiap akun.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, terdapat beberapa permasalahan yang menjadi fokus dalam penelitian ini, di antaranya :

1. Bagaimana cara mengidentifikasi akun *buzzer* di *TikTok* berdasarkan kemiripan komentar menggunakan metode *Cosine Similarity*?
2. Bagaimana performa algoritma *Random Forest*, *SVM*, *Decision Tree*, dan *Logistic Regresion* dalam mendeteksi akun *buzzer* di *TikTok* berdasarkan evaluasi model?
3. Bagaimana penerapan *cross-validation* dapat digunakan untuk mengevaluasi kestabilan dan konsistensi model dalam mendeteksi akun *buzzer* di *TikTok*, serta memastikan tidak adanya *overfitting*?

## 1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan untuk menjaga fokus pada tujuan utama. Batasan-batasan tersebut adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini terbatas pada komentar *TikTok* yang dikumpulkan dalam bentuk file CSV. Penelitian ini hanya fokus pada teks komentar, tanpa mempertimbangkan data lain seperti video, audio, atau interaksi pengguna seperti *like* dan *share*.
2. Penelitian ini menggunakan metode *machine learning* dengan pendekatan *TF-IDF* dan *Cosine Similarity* untuk menganalisis kemiripan antar-komentar. Algoritma *K-means* digunakan untuk melakukan *clustering*, sedangkan untuk pemodelan digunakan *Random Forest*, *SVM*, *Decision Tree*, dan *Regresi Logistik*. Metode lain seperti analisis sentiment, topic modeling atau model deep learning tidak termasuk dalam penelitian ini.
3. Variabel yang dianalisis dalam penelitian ini meliputi nilai *Cosine Similarity* antar-komentar (*max\_cosine\_sim*), jumlah komentar serupa (*similar\_comments\_count*), jumlah total komentar (*comment\_count*), serta selisih waktu antar-komentar (*time\_diff*). Variabel lain seperti jumlah *like*, jumlah pengikut, atau informasi profil pengguna tidak diperhitungkan dalam penelitian ini.

#### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengidentifikasi akun *buzzer* di *TikTok* dengan menganalisis kemiripan komentar menggunakan metode *TF-IDF* dan *Cosine Similarity*, serta mengelompokkan komentar yang memiliki pola serupa menggunakan algoritma *K-means* untuk mendeteksi aktivitas *buzzer* dengan lebih akurat. Selain itu, penelitian ini bertujuan untuk mengevaluasi kinerja model *machine learning* seperti *Random Forest*, *SVM*, *Decision Tree*, dan *Logistic Regression* dalam klasifikasi akun *buzzer*, serta memastikan model tidak mengalami *overfitting* melalui teknik validasi silang (*cross-validation*). Penelitian ini juga akan menghasilkan sebuah aplikasi berbasis web yang memungkinkan pengguna untuk mengunggah data komentar *TikTok* dalam format CSV, melakukan analisis, mendapatkan hasil deteksi *buzzer* secara otomatis, serta membangun model dengan algoritma yang digunakan dalam penelitian ini.

#### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis. Dari segi teoritis, penelitian ini memperkaya literatur tentang identifikasi *buzzer* di media sosial, khususnya di *TikTok*, menggunakan metode *machine learning*, *TF-IDF*, *Cosine Similarity*, dan algoritma *K-means*. Hasil penelitian ini juga dapat menjadi referensi bagi penelitian selanjutnya yang berfokus pada analisis perilaku akun *buzzer* atau studi mengenai disinformasi di platform media sosial.

Secara praktis, penelitian ini bermanfaat bagi pengguna umum, organisasi, dan peneliti lainnya. Aplikasi berbasis web yang dihasilkan dari penelitian ini dapat membantu pengguna mendeteksi akun *buzzer* secara otomatis dengan mengunggah data komentar *TikTok*, sehingga mereka bisa menganalisis interaksi publik dengan lebih efektif. Bagi organisasi atau perusahaan, aplikasi ini dapat digunakan untuk memonitor aktivitas *buzzer* dalam kepentingan bisnis, pemasaran, atau politik. Selain itu, penelitian ini dapat menjadi landasan bagi peneliti yang ingin mengembangkan metode lebih lanjut atau menerapkannya di platform lain.

## 1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun dalam lima bab utama yang berisi uraian secara garis besar mengenai isi dari masing-masing bab, sebagai berikut :

1. **BAB I PENDAHULUAN** : Bab ini berisi latar belakang masalah, rumusan masalah, tujuan penelitian, batasan masalah, dan manfaat penelitian. Bab ini juga menjelaskan urgensi dari penelitian yang dilakukan, serta bagaimana penelitian ini akan memberikan kontribusi terhadap studi yang ada.
2. **BAB II TINJAUAN PUSTAKA** : Bab ini menguraikan tinjauan pustaka yang mendukung penelitian, termasuk dasar-dasar teori yang digunakan, penelitian-penelitian terdahulu yang relevan, serta penjelasan tentang metode *machine learning* yang diterapkan dalam penelitian ini. Di dalamnya juga terdapat pembahasan terkait algoritma yang digunakan seperti *TF-IDF*, *Cosine Similarity*, *K-means*, dan model klasifikasi seperti *Random Forest*, *SVM*, *Decision Tree*, dan *Logistic Regression*.
3. **BAB III METODE PENELITIAN** : Bab ini menjelaskan metode penelitian yang digunakan dalam proses penelitian ini. Terdapat uraian tentang objek penelitian, teknik pengumpulan data, proses preprocessing data, serta tahap-tahap pemodelan dan evaluasi model menggunakan teknik validasi silang (*cross-validation*). Penjelasan mengenai pengembangan aplikasi berbasis web juga disertakan di bab ini.
4. **BAB IV HASIL DAN PEMBAHASAN** : Bab ini berisi hasil penelitian, termasuk hasil analisis data, evaluasi performa model menggunakan metrik seperti *akurasi*, *presisi*, *recall*, dan *F1-score*. Selain itu, bab ini juga menjelaskan hasil pengelompokan komentar dengan algoritma *K-means* dan bagaimana hasil deteksi *buzzer* divisualisasikan di aplikasi web yang dikembangkan. Bab ini juga membahas temuan dan analisis lebih lanjut dari

penelitian yang dilakukan.

5. **BAB V PENUTUP** : Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, termasuk jawaban atas rumusan masalah yang diajukan. Selain itu, bab ini memberikan saran-saran yang dapat menjadi rekomendasi untuk pengembangan penelitian lebih lanjut, baik dalam penerapan model maupun pengembangan aplikasi web.

