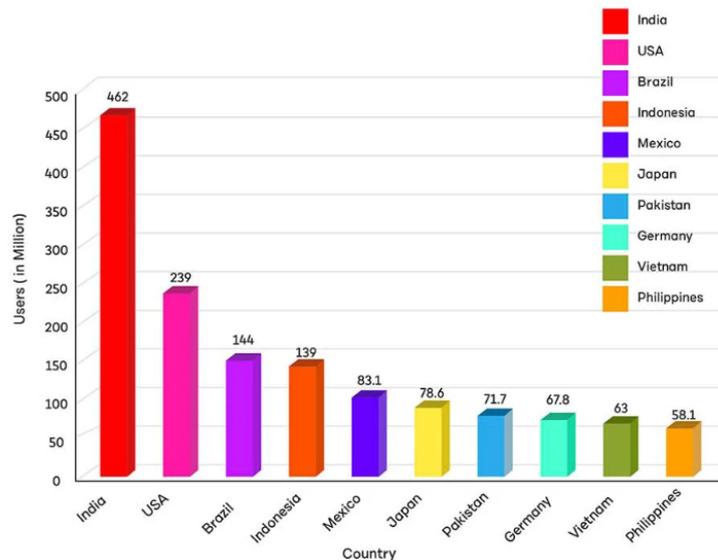


BAB I

PENDAHULUAN

1.1 Latar Belakang

YouTube adalah platform berbagi video yang menampilkan beragam jenis konten dari para kreator, dengan variasi genre dan durasi yang bervariasi[1]. Youtube menjadi salah satu platform di dunia dengan lebih dari 95% populasi internet yang menggunakannya. Secara global jumlah pengguna Youtube mencapai 2,70 miliar pada tahun 2024, sedangkan di Indonesia pengguna Youtube mencapai 139 juta dan termasuk 10 besar pengguna Youtube paling banyak didunia [2]. Untuk melihat tampilan grafik pengguna Youtube terbanyak menurut negara terdapat pada Gambar 1.1



Gambar 1. 1 Grafik Pengguna Youtube Menurut Negara Tahun 2024

Berdasarkan Gambar 1.1 di atas negara Indonesia menduduki peringkat keempat dengan jumlah pengguna Youtube terbanyak. Negara lain yang menduduki peringkat atas antara lain India 462 juta, USA 239 juta, Brazil 144 juta, Mexico 83,1 juta, Japan 78,6 juta, Pakistan 71,7 juta, Germany 67,8 juta, Vietnam 63 juta dan Philippines 58,1 juta. Selain itu ada

beberapa *genre* Youtube yang sedang populer seperti *podcast*, *asmr*, *fashion*, *vlog* harian, teknologi, *game*, politik dan lain sebagainya [3]. Dari salah satu *genre* tersebut yang sedang populer di Indonesia yaitu genre politik. Biasanya memiliki konten-konten seperti debat dan diskusi politik, komedi politik, berita politik dan edukasi politik. Salah satu kanal Youtube yang menciptakan diskusi mengenai politik yaitu Dirty Vote.

Kanal YouTube Dirty Vote merupakan kanal yang dirilis pada 9 Februari 2024 dan langsung menarik perhatian publik karena mengangkat isu-isu penting terkait politik di Indonesia. Video unggahan perdananya berjudul “*Dirty Vote – full movie*” yang diunggah pada tanggal 11 februari 2024, menghadirkan tiga pakar di bidang hukum tata negara yaitu Feri Amsari, Zainal Arifin Mochtar dan Bivitri Susanti. Ketiganya memberikan penjelasan mendalam mengenai kondisi demokrasi di Indonesia yang dinilai semakin merosot, terutama menjelang Pemilu 2024. Video ini memaparkan pihak-pihak yang memiliki kekuasaan melakukan berbagai dugaan kecurangan, mulai dari rekayasa hukum hingga intervensi terhadap lembaga negara. Penyajian informasi disampaikan secara lugas dan edukatif, dengan gaya narasi mirip film dokumenter. Tayangan ini mengajak masyarakat untuk lebih kritis dan menyadari pentingnya menjaga prinsip demokrasi yang sehat.

Permasalahan utama yang diangkat dalam video ini berfokus pada kondisi demokrasi Indonesia yang semakin mengkhawatirkan. Demokrasi seharusnya menjamin keadilan dan kebebasan dalam pemilu, namun yang terjadi justru sebaliknya terdapat indikasi kuat adanya penyalahgunaan kekuasaan untuk kepentingan politik tertentu. Hal ini menimbulkan keresahan di tengah masyarakat serta mendorong lahirnya berbagai opini publik, khususnya melalui media sosial seperti YouTube. Fokus pada opini publik penting untuk memahami persepsi masyarakat terhadap isu-isu politik yang sensitif. Dengan mengklasifikasikan komentar ke dalam kategori positif dan negatif arah sentimen publik bisa dianalisis secara lebih terstruktur.

Analisis sentimen merupakan proses pengolahan teks yang bertujuan untuk mendeteksi opini atau emosi seseorang yang terkandung dalam suatu

informasi, khususnya yang bersifat subjektif[4]. Analisis sentimen mengekstraksi opini atau pendapat dari sebuah dokumen, media sosial, *review blog*, serta data-data lainnya[5]. Dalam penerapannya, analisis sentimen tidak lepas dari tantangan teknis, khususnya dalam bidang *Natural Language Processing* (NLP). Tantangan tersebut meliputi penanganan bahasa informal, penggunaan bahasa campuran seperti campuran bahasa Indonesia dan Inggris, makna ganda dalam kata atau kalimat, penggunaan sarkasme, serta variasi gaya bahasa antar pengguna. Komentar di media sosial sering kali tidak terstruktur, mengandung singkatan, emoji, bahkan kesalahan penulisan, yang membuat proses pemrosesan bahasa alami menjadi semakin kompleks[6]. Untuk mengatasi permasalahan tersebut, diperlukan metode klasifikasi yang tidak hanya efisien secara komputasi tetapi juga mampu memberikan performa yang baik meskipun data teks bersifat tidak rapi atau bervariasi. Salah satu algoritma yang sering dimanfaatkan dalam menyelesaikan permasalahan klasifikasi teks yaitu *Naive Bayes Classifier*.

Naive Bayes Classifier adalah algoritma klasifikasi yang mengandalkan pendekatan probabilistik sederhana, berlandaskan pada frekuensi kemunculan serta kombinasi nilai dalam dataset [7]. Algoritma ini memiliki beberapa keunggulan, seperti kecepatan dalam perhitungan, kemampuan menghasilkan akurasi tinggi meski algoritmanya sederhana, serta mampu menangani data yang tidak lengkap[8]. Tingkat akurasi yang diperoleh dipengaruhi oleh kualitas data latih yang digunakan, dan akurasi cenderung optimal apabila data latih mencerminkan keseluruhan variasi dari data yang tersedia [9].

Selain mampu menghadapi tantangan dalam analisis teks, pemilihan *Naive Bayes Classifier* juga didukung oleh berbagai studi sebelumnya yang menunjukkan kinerjanya yang kompetitif dibandingkan metode lain. Beberapa penelitian yang membandingkan performa algoritma klasifikasi bahwa pendekatan *Naive Bayes* menunjukkan performa yang unggul dalam melakukan analisis sentimen. Penelitian Trifebi Shina Sabrila, Mujaddid Izzul Fikri, dan Yufis Azhar algoritma *Naive Bayes* menunjukkan hasil

tingkat akurasi yang lebih tinggi, yakni sebesar 73,65%, jika dibandingkan dengan metode *Support Vector Machine* [10]. Pada studi lain, Muhammad Yasir dan Robertus Suraji menunjukkan bahwa algoritma *Naïve Bayes* mendapatkan tingkat akurasi sebesar 90%, lebih lebih daripada *Decision Tree* (83%) dan *Random Forest* (87%)[11]. Penelitian yang dilakukan oleh Muhammad Syarifuddin juga mendukung keunggulan *Naïve Bayes*, menunjukkan akurasi sebesar 63,21%, sementara algoritma *K-Nearest Neighbor* hanya mencapai 58,10% [12]. Sementara itu, dalam studi yang dilakukan oleh Puji Winar Cahyo dan Ulfi Saidata Aesy, meskipun algoritma SVM mencatatkan akurasi tertinggi sebesar 0,74, *Naïve Bayes* tetap menunjukkan kinerja kompetitif dengan akurasi 0,62, melampaui model LSTM yang hanya memperoleh akurasi 0,47 [13].

Berdasarkan berbagai hasil yang telah dicapai pada penelitian sebelumnya, peneliti memutuskan untuk menggunakan algoritma *Naïve Bayes Classifier* dalam studi ini. Algoritma ini dipilih karena terbukti mampu memberikan performa yang unggul dalam menyelesaikan tugas klasifikasi teks, mudah diimplementasikan, serta mampu menangani tantangan yang muncul dalam analisis komentar di platform media sosial secara efektif. Tujuan penelitian ini adalah untuk mengklasifikasi sentimen masyarakat terhadap kanal YouTube Dirty Vote serta menguji kinerja algoritma *Naïve Bayes Classifier*.

1.2 Rumusan Masalah

Berdasarkan penjelasan mengenai latar belakang yang telah disampaikan, perumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana distribusi klasifikasi komentar sentimen positif dan negatif terhadap video “*Dirty Vote – Full Movie*” pada kanal YouTube Dirty Vote dengan penerapan algoritma *Naïve Bayes Classifier*?
2. Bagaimana kinerja algoritma *Naïve Bayes Classifier* dalam mengidentifikasi sentimen dari komentar video tersebut?

1.3 Tujuan dan Manfaat Penelitian

Penelitian ini bertujuan untuk mencapai sejumlah sasaran yang berhubungan dengan analisis sentimen pada komentar video “*Dirty Vote – Full Movie*” yang diunggah di kanal YouTube Dirty Vote. Berikut adalah tujuan dari penelitian ini :

1. Mengidentifikasi distribusi klasifikasi komentar sentimen positif dan negatif terhadap video “*Dirty Vote – Full Movie*” dengan memanfaatkan algoritma *Naïve Bayes Classifier*.
2. Menganalisis tingkat akurasi serta mengevaluasi kinerja dan efektivitas dari algoritma *Naïve Bayes Classifier* digunakan untuk mengklasifikasikan sentimen dari komentar-komentar video tersebut.

Adapun manfaat yang ingin dicapai melalui penelitian ini antara lain :

1. Memberikan evaluasi kepada pembuat konten mengenai tanggapan publik, khususnya terkait isu politik.
2. Menambah wawasan ilmiah tentang analisis sentimen di media sosial dan meningkatkan kesadaran masyarakat tentang pentingnya analisis kritis terhadap konten politik.

1.4 Batasan Masalah Penelitian

Batasan-batasan terakut masalah dalam penelitian ini adalah sebagai berikut:

1. Sumber data yang digunakan dari video salah satu kanal Dirty Vote yang berjudul “*Dirty Vote - full movie*”, serta data diambil pada tanggal 20 Maret 2024.
2. Hasil klasifikasi berupa sentiment positif dan negatif.
3. Pengolahan data dan implementasi analisis sentimen menggunakan bahasa pemrograman *python*.
4. Dataset diklasifikasikan menjadi data pelatihan dan data pengujian dengan proporsi 90 banding 10.

1.5 Metode Penelitian

Langkah-langkah yang dilakukan penelitian ini mencakup tinjauan pustaka, perumusan masalah, pengumpulan data, *preprocessing* data, pelabelan, visualisasi, pembobotan *term*, klasifikasi, evaluasi, hingga penyusunan kesimpulan.

Studi literatur dilakukan dengan mengkaji jurnal dan artikel ilmiah dari sumber terpercaya dalam lima tahun terakhir. Perumusan masalah berdasarkan berita dan artikel terkait kanal Dirty Vote. Data dikumpulkan dari komentar video YouTube “*Dirty Vote – full movie*” menggunakan YouTube API v3 dengan total 63 ribu komentar.

Pada tahap berikutnya, setelah *preprocessing* data seperti *case folding*, tokenisasi, *stopword removal*, *cleaning*, normalisasi dan *stemming*, data dilabeli dengan menggunakan *library* TextBlob untuk mengkategorikannya menjadi sentimen positif dan negatif. Selanjutnya visualisasi dari hasil pelabelan ini ditampilkan dalam bentuk *wordcloud* dan grafik.

Sebelum dilakukan fitur ekstraksi, data dibedakan menjadi dua kategori yaitu data pelatihan dan data pengujian. Fitur-fitur yang relevan diekstraksi dengan metode TF-IDF, untuk meningkatkan akurasi pada proses klasifikasi metode *Naive Bayes Classifier*. Kinerja model dinilai menggunakan *confusion matrix* serta metrik pengukuran seperti akurasi, *recall*, presisi, dan *f1-score*. Tahap akhir adalah menyusun kesimpulan berdasarkan hasil analisis serta saran untuk pengembangan peneliti selanjutnya.