

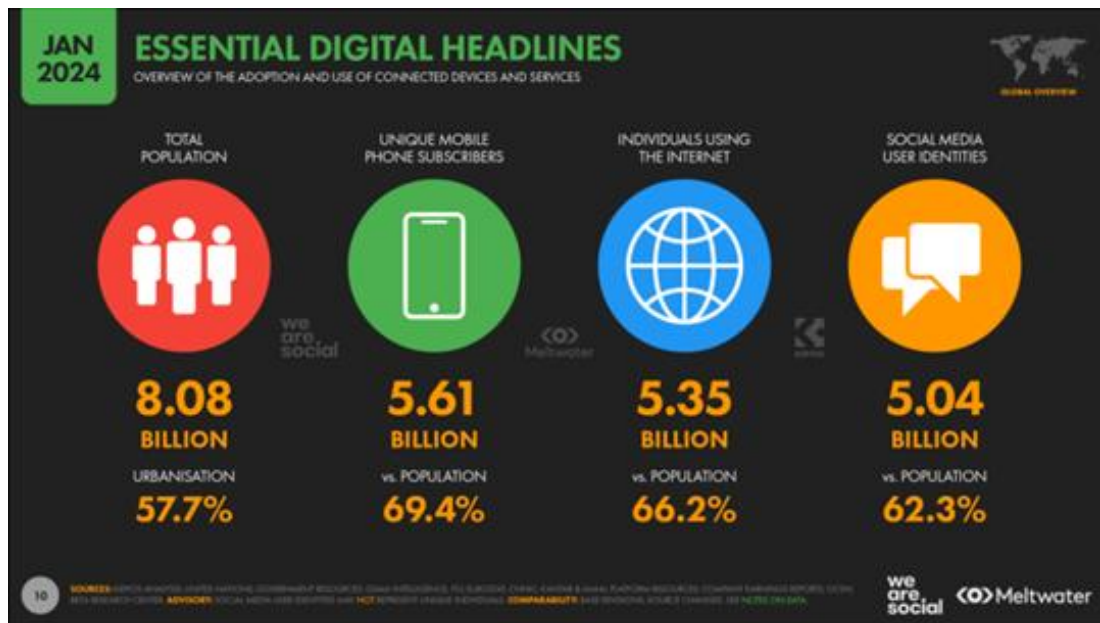
# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Teknologi Informasi dan Komunikasi saat ini telah banyak berkembang seiring berjalannya waktu. Pesatnya perkembangan TIK menjadikan internet sebagai alat komunikasi utama yang sangat diminati oleh Masyarakat [1]. Semakin tinggi teknologi informasi dan komunikasi yang ditemukan dan digunakan maka tidak terlepas dari bagaimana perkembangan komunikasi di masyarakat dan bagaimana masyarakat sosial tersebut berinteraksi [2]. Hal tersebut membuat hampir semua orang memiliki perangkat komunikasi yang memungkinkan untuk berkomunikasi dengan semua orang diseluruh dunia melalui media sosial. Dari berbagai kalangan mulai remaja hingga dewasa, hampir semua masyarakat Indonesia memiliki dan menggunakan media sosial sebagai salah satu sarana guna memperoleh dan menyampaikan informasi ke publik [3].

Menurut McGraw Hill Dictionary, Media sosial adalah sarana yang digunakan oleh orang-orang untuk berinteraksi satu sama lain dengan cara menciptakan, berbagi, serta bertukar informasi dan gagasan dalam sebuah jaringan dan komunitas virtual [4]. Menurut data dari WeAreAsocial seperti pada Gambar 1.1, Saat ini pengguna media sosial di dunia mencapai lebih dari 5 miliar jiwa dari total 8,08 miliar jiwa populasi di Dunia. Indonesia sendiri memiliki 139 juta pengguna media sosial pada Januari 2024, setara dengan 49,9 persen dari total populasi [5]. Hal tersebut menunjukkan besarnya minat masyarakat indonesia pada media sosial di jaman yang modern ini. Terdapat berbagai macam media sosial yang banyak digunakan di Indonesia, yaitu Instagram, Facebook, Tiktok, X, bahkan Youtube.



**Gambar 1.1 Populasi dan Pengguna Media Sosial di Dunia**

Banyaknya media sosial dan pengguna media sosial, tentunya memunculkan dampak negatif dan positif bagi penggunanya. Salah satu dampak negatif dari media sosial adalah interaksi secara tatap muka cenderung menurun dan membuat orang-orang menjadi kecanduan terhadap internet. Sedangkan Dampak positif dari media sosial adalah dapat memudahkan masyarakat untuk berinteraksi satu sama lain, memperluas pergaulan, jarak dan waktu bukan lagi masalah, penyebaran informasi dapat berlangsung secara cepat, lebih mudah dalam mengekspresikan diri dan memberikan pendapat [3].

Youtube merupakan salah satu aplikasi berbagi video yang paling sering dikunjungi di Indonesia. Menurut studi [6]. YouTube menyumbang 20% dari lalu lintas Web dan 10% dari total lalu lintas Internet. Di Indonesia, YouTube telah berkembang dengan pesat. Banyak saluran YouTube yang menyajikan berbagai jenis konten video yang diunggah oleh pengguna. YouTube adalah cara tercepat untuk menyebarkan informasi melalui video. YouTube menyediakan mekanisme interaksi sosial untuk menilai pendapat pengguna dan pandangan tentang video dengan cara voting, favorit, berbagi (share), serta komentar[7]. Data dari komentar pengguna yang melihat video Youtube ini berguna dalam mempelajari perilaku pengguna dan masyarakat. Dengan begitu banyaknya penggunaan youtube di seluruh dunia, pengaruhnya terhadap opini publik tidak bisa diabaikan.

Salah satu pembahasan/topik yang sedang ramai di Youtube Indonesia saat ini ialah topik mengenai Pilkada (Pemilihan Kepala Daerah) 2024 yang dilaksanakan

secara serentak. Tidak heran mengapa topik tersebut sedang ramai diperbincangkan, karena Indonesia merupakan negara demokrasi, sehingga pelaksanaan Pilkada (Pemilihan Kepala Daerah) ialah salah satu syarat penting bagi terciptanya sebuah negara yang demokratis [8]. Salah satu Pesta Demokrasi di suatu wilayah/daerah Indonesia diwujudkan dalam bentuk PILKADA (Pemilihan Kepala Daerah) untuk pemilihan Gubernur, Bupati, dan Walikota[8]. Banyak sekali isu-isu yang muncul menjelang Pilkada 2024 yang dilaksanakan secara demokratis. Seperti informasi mengenai siapa yang akan menjadi kontestan dalam Pilkada 2024 hingga adanya isu mengenai perubahan peraturan. Isu-isu tersebut banyak sekali memunculkan pro dan kontra sehingga masyarakat Indonesia beramai-ramai mengawal pilkada ini dengan memberikan tanggapan serta komentar melalui video yang diberitakan oleh akun-akun berita seperti Kompas.com, CNN, DLL. Komentar-komentar tersebut dapat memberikan gambaran mengenai sentimen atau perasaan masyarakat mengenai pilkada di Indonesia.

Klasifikasi sentimen, yang disebut juga dengan opinion mining, merupakan salah satu cabang ilmu dari data mining yang bertujuan untuk menganalisis, memahami, mengolah, dan mengekstrak data tekstual yang berupa opini terhadap entitas seperti produk, servis, organisasi, individu, dan topik tertentu[9]. Klasifikasi sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif [10]. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapat yang dikemukakan bersifat positif atau negatif. Namun seringkali sentimen pada sebuah komentar tidak dapat dilakukan dengan benar ketika komentar tersebut mengandung sarkasme.

Sarkasme merupakan suatu bentuk ironi khusus yang terjadi ketika seseorang menyampaikan informasi yang tersirat, biasanya memiliki makna yang berlawanan dengan apa yang diucapkan [11]. Kalimat sindiran atau sarkasme masih sering digunakan oleh kalangan publik untuk mengungkapkan maksud isi hati dan pikiran baik itu yang disampaikan secara langsung maupun tidak langsung [12]. Salah satu jenis sarkasme yang digunakan adalah jenis sarkasme yang berlainan makna yaitu kalimat sarkasme berbentuk kalimat positif yang memiliki makna negatif. Opini sarkasme ini sering sekali disampaikan seseorang secara tidak langsung melalui media sosial sehingga merupakan salah satu masalah yang mempengaruhi hasil analisis sentimen karena salah diklasifikasikan[12].

Untuk mengatasi permasalahan sarkasme dalam mengklasifikasikan sentiment, ada beberapa penelitian yang telah dilakukan. Salah satunya ada penelitian yang ditulis oleh [13] yang berjudul “Sentiment Analysis with Sarcasm Detection on Politician’s Instagram”. Penelitian ini menggunakan algoritma Naïve Bayes dan Random Forest untuk analisis sentimen serta analisis sentiment dengan deteksi sarkasme. Hasil menunjukkan bahwa akurasi terbaik pada model analisis sentimen diperoleh dengan menggunakan Random Forest dengan akurasi 72% dan akurasi model sarkasme dengan Random Forest adalah 71%. Selain itu berdasarkan literatur review yang ditulis oleh [14] yang berjudul “Sarcasm detection using machine learning algorithms in Twitter: A systematic review”. Literatur review ini membandingkan 31 penelitian mengenai deteksi sarkasme menggunakan algoritma machine learning dan hasil menunjukkan bahwa Support Vector Machine (SVM) merupakan AMLA terbaik dan paling umum digunakan untuk mendeteksi sarkasme.

Berdasarkan kedua penelitian tersebut, maka dilakukan penyusunan skripsi dengan judul “Komparasi Kinerja Algoritma SVM Dan RF dalam Klasifikasi Sentimen dengan Deteksi Sarkasme pada Komentar Youtube” dengan menggunakan algoritma Support Vector Machine dan Random Forest. Skripsi ini bertujuan untuk mengetahui hasil komparasi kinerja algoritma SVM dan RF dalam klasifikasi sentimen dengan deteksi sarkasme pada komentar video youtube terkait topik Pilkada 2024. Random Forest merupakan suatu metode ensemble yang terdiri dari beberapa decision tree sebagai pengklasifikasian[14]. Random Forest cocok untuk dataset biner karena menggunakan decision tree sebagai basis pembelajar dan cukup baik untuk mengklasifikasikan data dengan tipe biner. Sedangkan Support Vector Machine merupakan salah satu algoritma terbaik dan paling umum digunakan untuk mendeteksi sarkasme karena algoritma SVM mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda.

Skripsi ini dimulai dengan pengumpulan data dari fenomena yang terjadi kemudian dilakukan proses pengolahan data, pelabelan data, klasifikasi, dan diakhiri dengan pengujian serta pengimplementasian model pada website yang akan dirancang. Proses klasifikasi akan membandingkan dua algoritma machine learning SVM dan RF yang digunakan untuk mendapatkan hubungan semantis teks yang terbaik. Hasil evaluasi model akan melibatkan metrik utama seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score* yang memberikan penilaian kuantitatif terhadap kemampuan model dalam mengklasifikasikan sentimen dan sarkasme dengan tepat.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang dipaparkan sebelumnya maka dapat dirumuskan permasalahan, yaitu bagaimana hasil komparasi kinerja algoritma SVM dan RF dalam klasifikasi sentimen dengan deteksi sarkasme pada komentar video youtube terkait topik Pilkada 2024?

## 1.3. Batasan Masalah

Batasan masalah yang digunakan sebagai acuan dalam skripsi ini adalah sebagai berikut:

1. Data yang digunakan dalam skripsi ini diambil dari komentar video Youtube menggunakan bantuan *library* pendukung terkait Pilkada (Pemilihan Kepala Daerah) 2024 dari beberapa channel youtube yang terpercaya di antaranya, Kompas Tv, 7Metro Tv, Liputan 6, iNews, TvOne, Kumparan.
2. Data sentimen diklasifikasikan ke dalam dua kelas yaitu, sentimen negatif dan sentimen positif
3. Rentang waktu pengambilan data yang digunakan dalam skripsi ini yaitu mulai dari tanggal 1 Mei 2024 hingga 7 Oktober 2024.
4. Skripsi ini menggunakan pembobotan kata TF-IDF dan TF-ABS

## 1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan sebelumnya maka skripsi ini dibuat dengan tujuan, yakni untuk mengetahui hasil komparasi kinerja algoritma SVM dan RF dalam klasifikasi sentimen dengan deteksi sarkasme pada komentar video youtube terkait topik Pilkada 2024

## 1.5. Sistematika Penulisan

Sistematika penulisan skripsi ini bertujuan untuk membantu penyusunan laporan skripsi agar sesuai dengan acuan yang telah ditentukan dan mencapai tujuan penulisan yang telah ditetapkan. Beberapa tahapan yang harus dilalui dalam proses penyusunan skripsi ini adalah sebagai berikut:

### **BAB I            PENDAHULUAN**

Bab ini menjelaskan gambaran umum tentang skripsi yang akan dilakukan, termasuk latar belakang skripsi, rumusan masalah 6 yang akan dipecahkan, batasan masalah yang ditetapkan, tujuan penelitian, dan sistematika penulisan yang akan digunakan.

## **BAB II            TINJAUAN PUSTAKA**

Bab ini berisi tentang daftar penelitian yang sudah pernah dilakukan sebelumnya yang bertujuan untuk menjadi dasar penelitian. Penelitian yang dibahas adalah penelitian yang memiliki keterkaitan dengan skripsi ini seperti Youtube, *text mining*, *preprocessing*, klasifikasi sentimen, deteksi sarkasme, *Support Vector Machine*, *Random Forest*, dll.

## **BAB III            METODOLOGI PENELITIAN**

Bab ini berisi tentang metode penelitian yang akan dibuat yaitu meliputi studi literatur, analisis kebutuhan, dan perancangan model dan visualisasi.

## **BAB IV            PENGUJIAN DAN ANALISA**

Bab ini akan menguraikan hasil penelitian yaitu mengenai proses dalam penyusunan skripsi

## **BAB V            PENUTUP**

Bab ini berisi mengenai Kesimpulan dan saran yang diberikan oleh peneliti berdasarkan hasil skripsi ini.

## **DAFTAR PUSTAKA**

Berisi mengenai literatur yang digunakan sebagai pedoman yang membantu pengerjaan skripsi.