

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Dunia digital semakin canggih sehingga perkembangan informasi lebih mudah didapatkan melalui portal atau sosial media. Sosial media merupakan salah satu tempat untuk menyampaikan opini masyarakat. Opini masyarakat dapat berupa pujian atau ujaran kebencian dan *hoax* yang dapat menimbulkan perdebatan disosial media. Salah satu media diskusi adalah dengan penggunaan Twitter [1]. Pada bulan Juli 2023, Twitter mengalami perubahan nama menjadi X setelah diakuisisi oleh perusahaan Amerika Serikat, X Corp. Perubahan nama ini dilakukan dengan tujuan untuk memperluas fungsi X menjadi *superapp* [2].

X adalah salah satu sosial media terpopuler yang berperan sebagai wadah komunikasi di masyarakat. Salah satu layanan yang disediakan oleh X kepada penggunanya adalah pembuatan pesan status atau disebut "*tweet*" yang dapat dibaca oleh pengguna X lainnya dan biasa berisi ungkapan pendapat pengguna dalam berbagai topik dengan batasan sebanyak 140 karakter, namun setelah berganti nama X mampu membagi konten digital apapun tanpa dibatasi karakter [2] [3]. Kemudahan yang diberikan dalam mengakses sosial media dapat memberikan banyak opini yang disampaikan oleh masyarakat yang dapat memengaruhi stigma dan stereotip yang melekat pada seorang kandidat. Hal ini relevan dalam konteks sengketa pemilihan Presiden Indonesia 2024 [4] [5].

Pemilihan Umum (Pemilu) adalah proses yang bertujuan untuk memilih individu yang akan menduduki posisi politik tertentu, yang harus mengutamakan prinsip-prinsip demokrasi seperti kejujuran dan keadilan, dengan mematuhi hukum yang berlaku. Namun dalam pelaksanaan pemilihan umum seringkali terjadi berbagai pelanggaran, baik administratif maupun kriminal, serta terjadinya sengketa dalam proses pemilihannya, termasuk sengketa pemilihan presiden [6] [7]. Oleh karena itu, akan muncul berbagai opini masyarakat mengenai sengketa pemilihan presiden Indonesia 2024 di sosial media X. Dalam pemanfaatan informasi agar dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan dibutuhkan analisis. Analisis politik tradisional mengalami kesulitan

dalam mengklasifikasikan opini publik di media sosial secara akurat karena keterbatasan kosakata umum yang tidak dapat menangkap nuansa dan konteks khusus dalam diskusi politik [8]. Analisis sentimen dapat digunakan untuk menunjukkan sentimen positif atau negatif dari opini yang disampaikan oleh masyarakat [4].

Analisis sentimen bertujuan untuk menganalisis opini, sentimen dan emosi yang terdapat pada dokumen atau data. Tugas dasar dari analisis sentimen adalah untuk mengelompokkan sifat dari teks yang ada di dalam kalimat maupun pendapat yang pada umumnya terbagi menjadi kelas yaitu positif dan negatif [9]. Namun, salah satu tantangan utama dalam analisis sentimen adalah memilih dan mengoptimalkan algoritma yang paling tepat untuk mengklasifikasikan teks secara akurat. Algoritma *Random Forest* adalah algoritma yang sering digunakan dalam mengklasifikasikan teks, yang terdiri dari beberapa pohon keputusan sebagai *classifier*. *Random forest* juga sering disebut dengan metode *ensemble learning* yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi dan mengurangi risiko *overfitting*. Metode ini dipilih karena memiliki keunggulan dalam meminimalkan kesalahan prediksi yang sering terjadi pada metode lain, terutama ketika menghadapi data yang kompleks dan bervariasi [10].

Beberapa penelitian telah melakukan perbandingan antara algoritma untuk menemukan metode yang paling efektif. Misalnya, penelitian oleh Christina Purnama Yanti dkk, dalam "Perbandingan Metode *K-NN* Dan Metode *Random Forest* Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia" menunjukkan bahwa *Random Forest* sedikit lebih unggul dengan akurasi 73,37%, dibandingkan dengan metode *K-NN* dengan akurasi 72,86% [11]. Studi lain oleh Dimas Triyana dkk, yang berjudul "Implementasi Metode *Ensemble Majority Vote* pada Algoritma *Naïve Bayes* dan *Random Forest* untuk Analisis Sentimen Twitter Harga Tiket Pesawat Domestik", menunjukkan bahwa *Random Forest* mencapai akurasi sebesar 85,71% dan 84,91%, lebih baik daripada *Naive Bayes* yang hanya mencapai 71,75% dan 70,31% [12]. Selain itu, penelitian yang menggunakan metode Support Vector Machine (SVM) untuk analisis sentimen terkait Pilpres 2024 pada platform Twitter menunjukkan performa yang bervariasi berdasarkan

kernel yang digunakan, seperti kernel linier, polynomial, RBF, dan sigmoid, dengan akurasi tertinggi ditemukan pada kernel linier [13]. Berdasarkan beberapa penelitian tersebut, terlihat bahwa algoritma *Random Forest* menunjukkan hasil akurasi yang lebih unggul dibandingkan beberapa algoritma lain.

## 1.2 Kajian Riset Terdahulu

Kajian riset terdahulu adalah upaya peneliti untuk mencari perbandingan dan selanjutnya untuk menemukan inspirasi baru untuk penelitian selanjutnya di samping itu kajian terdahulu membantu penelitian dapat memposisikan penelitian serta menunjukkan orsinalitas dari penelitian. Dalam tahap ini, penelitian akan diuraikan secara singkat sebagai bentuk memperkuat alasan mengapa penelitian ini dilakukan. Tabel 1.1 adalah referensi jurnal sejenis yang dilakukan beberapa peneliti sebelumnya.

Tabel 1. 1 Referensi penelitian.

Nama Peneliti	Tahun	Judul Penelitian
Ayu Sagita, Ahmad Faqih, Gifthera Dwilestari, Bambang Siswoyo, Denni Pratama	2023	Penerapan Metode <i>Random Forest</i> Dalam Menganalisis Sentimen Pengguna Aplikasi Capcut Di Google Play Store
Christina Purnama Yanti, Ni Wayan Eva Agustini, Ni Luh Wiwik Sri Rahayu Ginantra, Dewa Ayu Putri Wulandari	2023	Perbandingan Metode <i>K-NN</i> Dan Metode <i>Random Forest</i> Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia
Yin Min Tun, Myo Khaing	2023	<i>A large-scale sentiment analysis using political tweets</i>
Dimas Triyana, Muhammad Muharrom Al Haromainy Hendra Maulana	2024	Implementasi Metode <i>Ensemble Majority Vote</i> Pada Algoritma <i>Naive Bayes</i> Dan <i>Random Forest</i> Untuk Analisis Sentimen Twitter Harga Tiket Pesawat Domestik
Muhammad Rusdi Rahman, Ahmad Febri Diansyah, Hanafi	2024	<i>Sentiment Analysis on the Shopee Application on Playstore Using the Random Forest Classification Method</i>

Pada penelitian yang dilakukan Ayu Sagita [14], dkk mengenai Penerapan Metode *Random Forest* Dalam Menganalisis Sentimen Pengguna Aplikasi Capcut

Di Google Play Store. Metode *Random Forest* dipilih untuk mengidentifikasi dan mengelompokkan tanggapan pengguna ke dalam kategori positif dan negatif secara efisien. Dataset yang digunakan dalam penelitian ini mencakup 5000 ulasan, mencerminkan keragaman pendapat dari pengguna yang berpartisipasi aktif. Tahapan *pre processing* data melibatkan proses *cleansing*, *case folding*, tokenisasi, *stopword removal*, dan lemmatisasi untuk memastikan kualitas data yang baik sebelum dilakukan analisis sentimen. Selanjutnya, pembobotan kata dilakukan dengan metode TF-IDF untuk memberikan nilai bobot pada kata-kata yang mempengaruhi sentimen pengguna. Hasil penelitian menunjukkan bahwa metode *Random Forest* memberikan tingkat akurasi yang tinggi dalam menganalisis sentimen pengguna CapCut, dengan akurasi sebesar 86%, presisi 89%, *recall* 81%, dan *f1-score* 85%.

Pada penelitian yang dilakukan Christina Purnama Yanti [11], dkk mengenai Perbandingan Metode *K-NN* dan Metode *Random Forest* untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia. Penelitian ini menggunakan metode *K-NN* dan *Random Forest* dan tujuan dari penelitian ini untuk mengetahui perbandingan dari kedua metode tersebut dalam analisis sentimen isu minyak goreng. Dari dua metode yang digunakan masing-masing akan dibagi menjadi tiga skenario penelitian, yang pertama skenario 1 berjumlah 500 data, skenario 2 berjumlah 800 data, dan skenario 3 berjumlah 1.000 data, dimana untuk rasio data latih dan data uji yaitu 80:20. Untuk hasil pengujian metode *K-NN* pada skenario 2 lebih unggul dengan presentase akurasi sebesar 74,58%, presisi 56,75% dan *recall* 44,57% dan hasil yang paling rendah yaitu metode *K-NN* skenario 1 dengan presentase akurasi sebesar 71,50%, presisi 47,83% dan *recall* 37,45%. Untuk rata-rata hasil pengujian dari metode *K-NN* yaitu akurasi 72,86%, presisi 52,26% dan *recall* 41,04%. Sedangkan rata-rata hasil pada metode *random forest* yaitu akurasi 73,37%, presisi 52,26% dan *recall* 34,28%.

Pada penelitian yang dilakukan oleh Yin Min Tun dkk, [8] mengenai Analisis sentimen skala besar menggunakan *tweet* politik. Dalam sistem yang diusulkan, kosakata khusus domain politik dibangun dengan menggunakan pendekatan terawasi untuk mengekstrak kata-kata opini politik ekstrem dan fitur

dalam *tweet*. Sistem analisis sentimen multi-kelas politik (*Political Multi-Class Sentiment Analysis/PMSA*) pada *platform big data* dikembangkan untuk memprediksi kecenderungan tweet guna menyimpulkan hasil pemilu dengan melakukan analisis pada berbagai dataset politik: termasuk dataset pemilu Trump dan politik BBCNews. Analisis komparatif merupakan hasil eksperimen yang menunjukkan klasifikasi teks politik yang lebih baik dengan menggunakan tiga model yang berbeda (*multinomial naïve Bayes (MNB)*, *decision tree (DT)*, *linear support vector classification (SVC)*). Hasil penelitian menggunakan algoritma *Naïve Bayes Multinomial* menunjukkan akurasi sebesar 74%, algoritma *Random Forest* menunjukkan akurasi sebesar 96% dan algoritma SVC Linier menunjukkan akurasi sebesar 98%.

Pada penelitian yang dilakukan oleh Dimas Triyana [15] dkk, mengenai Implementasi Metode *Ensemble Majority Vote* Pada Algoritma *Naive Bayes* Dan *Random Forest* Untuk Analisis Sentimen Twitter Harga Tiket Pesawat Domestik. Fokus Penelitian ini menggunakan algoritma *Naive Bayes* dan *Random Forest* kemudian diimplementasikan pada *Ensemble Majority Vote*. Pada analisis sentimen pada tweet terkait harga tiket penerbangan domestik. Skenario pertama *Naive Bayes* menggunakan data *training* dan *testing* dengan rasio 80:20, menghasilkan akurasi sebesar 70,31% setelah menerapkan SMOTE untuk menangani ketidakseimbangan pada *labelling* data. Skenario kedua, dengan rasio 70:30, mendapat akurasi 71,75%. Metode *Random Forest* juga digunakan sebagai algoritma perbandingan. Pada skenario pertama, *Random Forest* mencapai akurasi 84,91%, dan pada skenario kedua, 85,71%. Penggabungan kedua algoritma menggunakan *ensemble majority voting* menghasilkan akurasi 85,88% dan 85,87% untuk kedua skenario, menunjukkan peningkatan akurasi prediksi dibandingkan model individu algoritma.

Pada penelitian yang dilakukan Muhammad Rusdi Rahman [16] dkk, mengenai *Sentiment Analysis on the Shopee Application on Playstore Using the Random Forest Classification Method*. Dalam menganalisis kepuasan pelanggan atau konsumen terhadap layanan perusahaan, sangat penting bagi perusahaan untuk menemukan kekurangan layanan dan mengetahui harapan pengguna terhadap

perusahaan. Penelitian ini bertujuan untuk membangun analisis sentimen pada aplikasi Shopee di Google Playstore. Metode yang digunakan mencakup TF-IDF sebagai vektorisasi teks, *Random Forest* sebagai model klasifikasi, dan *Evaluation Matrix* sebagai model evaluasi, dengan metrik akurasi, presisi, *recall*, dan *f1-Score*. Berdasarkan hasil penelitian ini, model yang kami gunakan mencapai tingkat akurasi sebesar 94%, presisi 91%, *recall* 91%, dan *f1-Score* sebesar 93%.

Perbedaan penelitian ini dengan penelitian yang lain adalah melakukan rancang bangun sistem analisis sentimen terhadap sengketa pemilihan Presiden Indonesia 2024 dengan menggunakan metode *Random Forest*. Selain itu, penelitian ini juga mencakup implementasi sistem dalam aplikasi web sederhana yang memungkinkan visualisasi sentimen, yang membedakannya dari penelitian serupa yang mungkin hanya terbatas pada analisis yang lebih umum atau tidak menggunakan *Random Forest* sebagai metode utama.

### **1.3 Rumusan Masalah**

Berdasarkan latar belakang di atas, maka dapat dirumuskan beberapa hal berikut ini:

1. Bagaimana analisis sentimen pengguna sosial media X terhadap sengketa pemilihan Presiden Indonesia 2024 menggunakan *Random Forest*?
2. Bagaimana kinerja model *Random Forest* dalam mengklasifikasikan sentimen positif dan negatif pada sosial media X terhadap sengketa pemilihan Presiden Indonesia 2024?

### **1.4 Tujuan**

Adapun Tujuan yang dilakukan pada penelitian ini adalah:

1. Menganalisis sentimen pada sosial media X terhadap sengketa pemilihan Presiden Indonesia 2024 menggunakan *Random Forest*.
2. Mengukur kinerja model *Random Forest* dalam mengklasifikasikan sentimen positif dan negatif pada sosial media X terhadap sengketa pemilihan Presiden Indonesia 2024.

### **1.5 Manfaat**

Adapun manfaat dari penelitian ini adalah:

1. Manfaat Akademik

Penelitian ini diharapkan dapat memberikan kontribusi akademik mengenai perkembangan di bidang *machine learning*, terutama tentang *Random Forest*.

## 2. Manfaat Praktis

Hasil dari penelitian ini diharapkan dapat membantu pemerintah dalam merumuskan dan menentukan kebijakan publik yang lebih baik dan membuat keputusan yang berdampak positif bagi masyarakat secara keseluruhan.

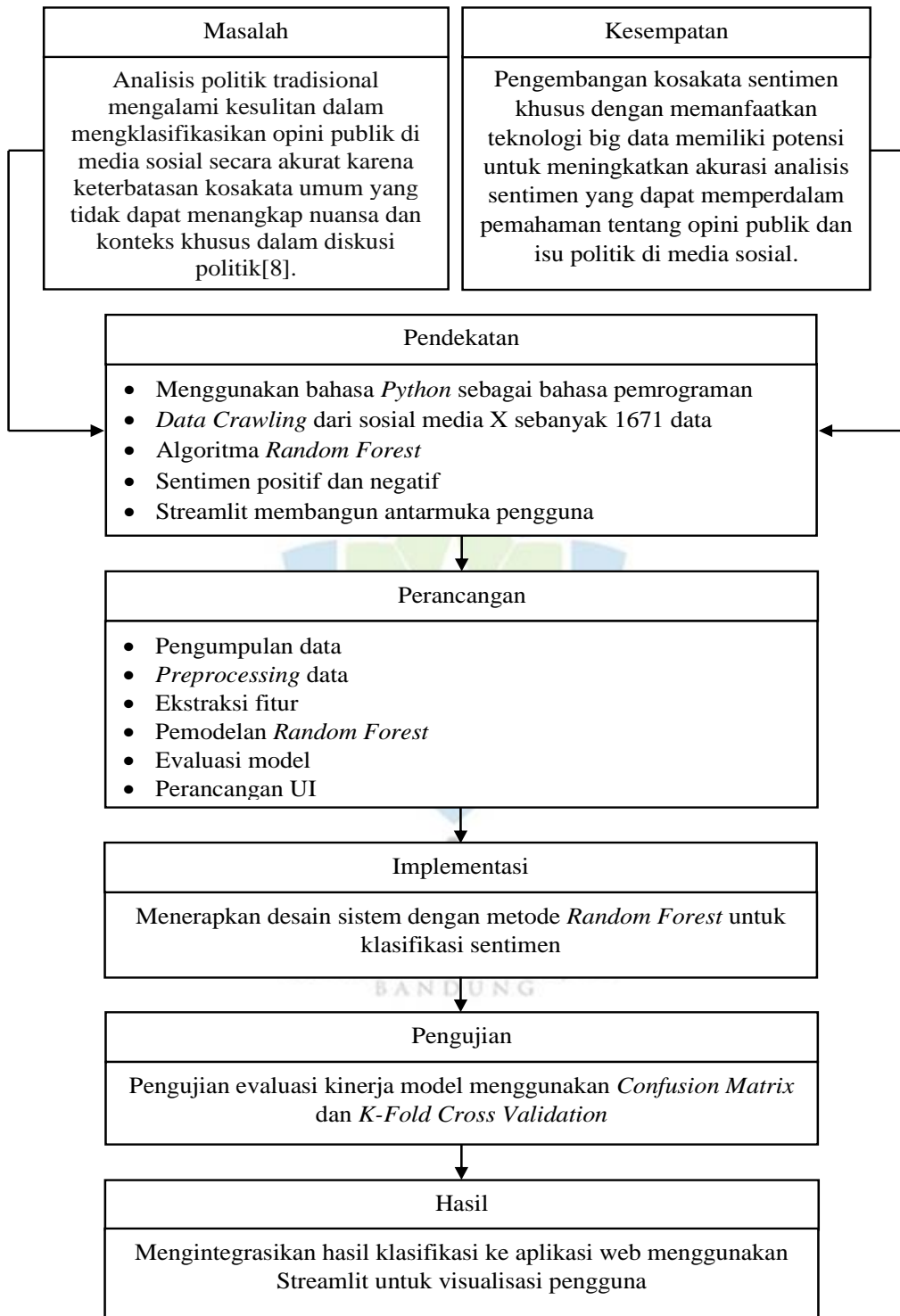
### 1.6 Batasan Masalah

Batasan masalah untuk menghindari adanya penyimpangan atau perluasan topik, agar penelitian lebih terarah dan mudah dibahas, sehingga dapat mencapai Tujuan penelitian. Beberapa batasan dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan diambil dari media sosial X pada bulan Februari 2024 – April 2024.
2. Sumber data yang digunakan diambil dari sosial media X.
3. Data *crawling* menggunakan kata kunci “Sengketa Pilpres”.
4. Penggunaan *tweet* bahasa Indonesia.
5. Tidak mengklasifikasikan akun palsu dan *Bot*.
6. Menggunakan metode *Random Forest*.
7. Hasil klasifikasi berupa sentimen positif dan negatif.
8. Aplikasi ini berfokus pada halaman *user interface* bukan pada pengembangan aplikasi.

### 1.7 Kerangka Berpikir

Kerangka berpikir ialah narasi atau pernyataan tentang kerangka konsep pada pemecahan masalah yang telah teridentifikasi atau dirumuskan. Kerangka berpikir dalam penelitian ini digambarkan seperti pada Gambar 1.1



Gambar 1. 1 Kerangka berpikir.



## **1.8 Sistematika Penulisan**

Sistematika penulisan merupakan suatu tahap penyusunan data dan penulisan dalam suatu laporan yang terdiri dari 6 bab agar dapat menghasilkan penulisan yang baik, diantaranya sebagai berikut:

### **BAB I PENDAHULUAN**

Pada bab ini diuraikan tentang latar belakang, kajian riset terdahulu, rumusan masalah, tujuan, manfaat, batasan masalah, kerangka pemikiran, dan sistematika penulisan.

### **BAB II TEORI DASAR**

Pada bab ini berisi tentang teori dasar yang digunakan dalam penelitian serta memberikan gambaran peralatan yang digunakan dalam penelitian ini.

### **BAB III METODOLOGI DAN JADWAL PENELITIAN**

Pada bab ini menjelaskan metode dan tahapan-tahapan yang dilakukan ketika melakukan penelitian rancang bangun sistem analisis sentimen pengguna sosial media X terhadap sengketa pemilihan Presiden Indonesia 2024 menggunakan *Random Forest*.

### **BAB IV PERANCANGAN DAN IMPLEMENTASI**

Bab ini menjelaskan rencana penelitian dengan merancang sistem analisis sentimen pengguna sosial media X terkait sengketa pemilihan presiden Indonesia 2024.

### **BAB V PENGUJIAN DAN ANALISIS**

Pada bab ini, dilakukan pengujian terhadap berbagai aspek sistem, termasuk akurasi model *Random Forest*, proses *preprocessing* data, serta kinerja keseluruhan sistem analisis sentimen.

### **BAB VI PENUTUP**

Bab ini menyimpulkan hasil penelitian secara keseluruhan dengan menarik kesimpulan berdasarkan tujuan yang telah ditetapkan. Selain itu, bab ini juga menyajikan saran untuk penelitian selanjutnya.